Pure Characteristics Demand Models and Distributionally Robust Mathematical Programs with Stochastic Complementarity Constraints

Jie Jiang · Xiaojun Chen

Abstract We formulate pure characteristics demand models under uncertainties of probability distributions as distributionally robust mathematical programs with stochastic complementarity constraints (DRMP-SCC). For any fixed first-stage variable and a random realization, the second-stage problem of DRMP-SCC is a monotone linear complementarity problem (LCP). To deal with ambiguity of probability distributions of the involved random variables in the stochastic LCP, we use the distributionally robust approach. Moreover, we propose an approximation problem with regularization and discretization to solve DRMP-SCC, which is a two-stage nonconvex-nonconcave minimax optimization problem. We prove the convergence of the approximation problem to DRMP-SCC regarding the optimal solution sets, optimal values and stationary points as the regularization parameter goes to zero and the sample size goes to infinity. Finally, preliminary numerical results for investigating distributional robustness of pure characteristics demand models are reported to illustrate the effectiveness and efficiency of our approaches.

Keywords distributionally robust · stochastic equilibrium · regularization · discrete approximation · pure characteristics demand

Mathematics Subject Classification (2010) 90C15 · 90C33 · 90C26

Jie Jiang
College of Mathematics and Statistics, Chongqing University, Chongqing, China
E-mail: jiangjiecq@163.com

Xiaojun Chen
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong
E-mail: maxjchen@polyu.edu.hk

1 Introduction

Pure characteristics demand models are widely used in microeconometrics to estimate parameters in utility functions of agents for given prices and production decisions [2,36]. Such models have advantages in inferring consumers' preference and behavior, but face computational challenges to solve the optimization problem with set-valued stochastic equilibrium constraints as follows:

$$
\begin{aligned}
&\min_{x \in X} \tfrac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle \\
&\text{s.t. } A_t \mathbb{E}_P[S_t(x,\xi)] \ni b_t \text{ for } t = 1, 2, \cdots, T,
\end{aligned}
\tag{1}
$$

where $X \subseteq \mathbb{R}^n$ is a compact and convex set, $H \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix, $c \in \mathbb{R}^n$, $\xi : \Omega \to \Xi \subseteq \mathbb{R}^\nu$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and supported on Ξ, $P = \mathbb{P} \circ \xi^{-1}$, $A_t \in \mathbb{R}^{l \times r}$, $b_t \in \mathbb{R}^l$, $S_t : \mathbb{R}^n \times \Xi \rightrightarrows \mathbb{R}^r$, $T \geq 1$ is the number of markets, $\mathbb{E}_P[S_t(x,\xi)]$ is a Aumann's (set-valued) expectation [1] for multifunction and for given $(x,\xi) \in X \times \Xi$,

$$
S_t(x,\xi) := \arg\max_s \{ \langle s, u_t(x,\xi) \rangle : \langle e, s \rangle \leq 1, s \geq 0 \}.
\tag{2}
$$

Here $u_t : \mathbb{R}^n \times \Xi \to \mathbb{R}^r$ is the consumers' utility function in market t and $e \in \mathbb{R}^r$ is the vector with all elements being 1. Obviously, $S_t(x,\xi)$ is bounded. Furthermore, we know from [32, Theorem 14.37] that $S_t(x,\xi)$ is closed-valued and measurable. Thus its Aumann's expectation is well-defined.

To efficiently solve problem (1), Pang, Su and Lee [29] characterized consumers' purchase decision in the constraints of problem (1) by linear complementarity problems and proposed the following quadratic program with stochastic complementarity constraints (QP-SCC):

$$
\begin{aligned}
&\min_{x \in X} \tfrac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle \\
&\text{s.t. } A_t \mathbb{E}_P[s_t(x,\xi)] = b_t \text{ for } t = 1, 2, \cdots, T, \\
&\qquad 0 \leq z_t(x,\xi) \perp M z_t(x,\xi) + q_t(x,\xi) \geq 0 \text{ for a.e. } \xi \in \Xi,
\end{aligned}
\tag{3}
$$

where

$$
M = \begin{pmatrix} 0 & e \\ -e^\top & 0 \end{pmatrix} \in \mathbb{R}^{(r+1) \times (r+1)}, \quad q_t(x,\xi) = \begin{pmatrix} -u_t(x,\xi) \\ 1 \end{pmatrix} \in \mathbb{R}^{r+1},
$$

$$
s_t(x,\xi) \in \mathbb{R}^r, \quad z_t(x,\xi) = \begin{pmatrix} s_t(x,\xi) \\ \gamma_t(x,\xi) \end{pmatrix} \in \mathbb{R}^{r+1}
$$

and a.e. is the short for almost everywhere. The pioneered QP-SCC formulation opened a way to develop optimization algorithms for solving pure characteristics demand models. In [4], Chen, Sun and Wets proposed a penalty approach:

$$
\begin{aligned}
&\min_{x \in X} \tfrac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle + \varrho \sum_{t=1}^T \|A_t \mathbb{E}_P[s_t(x,\xi)] - b_t\|^2 \\
&\text{s.t. } 0 \leq z_t(x,\xi) \perp M z_t(x,\xi) + q_t(x,\xi) \geq 0 \\
&\qquad \text{for a.e. } \xi \in \Xi \text{ and } t = 1, 2, \cdots, T,
\end{aligned}
\tag{4}
$$

where $\varrho > 0$ is a penalty parameter.

In problems (1)-(4), the probability distribution P is supposed to be known exactly. However, in practice the true probability distribution can hardly be acquired. This observation motivates us to use the distributionally robust approach [9,10,38,39] to introduce a class of distributionally robust stochastic mathematical programs with complementarity constraints (DRMP-SCC) as follows:

$$\min_{x,y} \Phi(x,y) := \theta(x) + \sup_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x,y(\xi),\xi)]) \tag{P}$$
$$\text{s.t. } x \in X, 0 \leq y(\xi) \perp M(\xi)y(\xi) + q(x,\xi) \geq 0 \text{ for a.e. } \xi \in \Xi,$$

where $y : \Xi \to \mathbb{R}^m$ is a measurable mapping, $\theta : \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}^l$, $h : \mathbb{R}^l \to \mathbb{R}$, $\mathcal{P} \subseteq \mathcal{P}(\Xi)$ is an ambiguity set with $\mathcal{P}(\Xi)$ denoting the collection of all probability distributions supported on Ξ, $q : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$, $M : \Xi \to \mathbb{R}^{m \times m}$ and $M(\xi)$ is positive semidefinite for a.e. $\xi \in \Xi$. For fixed x and ξ, a feasible vector $y(\xi)$ of problem (P) is a solution of the monotone LCP. The error bounds for the monotone LCP and exact penalty theory for mathematical programms with linear complementarity constraints established by Mangasarian et al. in [21–26] have inspired us to solve problem (P).

Throughout this paper, we assume that θ, h and $f(\cdot,\cdot,\xi)$ are continuously differentiable. Moreover, θ, h, M, q and $f(\cdot,\cdot,\xi)$ are Lipschitz continuous with Lipschitz moduli L_θ, L_h, L_M, L_q and $L_f(\xi)$ satisfying $\max_{P \in \mathcal{P}} \mathbb{E}_P[L_f^2(\xi)] < \infty$, respectively. We also assume that problem (P) satisfies the relatively complete recourse condition, that is, for every $x \in X$ and a.e. $\xi \in \Xi$, the solution set $\text{SOL}(M(\xi), q(x,\xi))$ of $\text{LCP}(M(\xi), q(x,\xi))$ in the constraints of problem (P) is nonempty.

Moreover, we assume that $\mathbb{E}_P[f(x,y(\xi),\xi)]$ is well-defined for all $P \in \mathcal{P}$ and $y(\xi) \in \text{SOL}(M(\xi), q(x,\xi))$, and the ambiguity set $\mathcal{P}$ is defined by a general moment form as follows:

$$\mathcal{P} = \{P \in \mathcal{P}(\Xi) : \mathbb{E}_P[\Psi(\xi)] \in \Gamma\}, \tag{5}$$

where Ψ is a continuous random mapping consisting of vectors and/or matrices with measurable random components, and Γ is a closed convex cone in the Cartesian product of some finite dimensional vector and/or matrix spaces, which is first considered in [40] and then in [18,45]. The ambiguity set defined in (5) is a very general form and includes many commonly-used moment ambiguity sets, such as the moment ambiguity set in [9]. For more examples, we refer to [18, Examples 3-5].

Let $\Xi^k = \{\xi^1, \cdots, \xi^k\}$ be a set of k samples of ξ and define the discrete approximation of $\mathcal{P}$ by

$$\mathcal{P}_k = \left\{p \in \mathbb{R}_+^k : \sum_{i=1}^k p_i = 1, \sum_{i=1}^k p_i \Psi(\xi^i) \in \Gamma\right\}.$$

We consider the discrete approximation problem of (P) as follows:

$$\min_{x,\mathbf{y}} \Phi_k(x,\mathbf{y}) := \theta(x) + \max_{p \in \mathcal{P}_k} h(F(x,\mathbf{y})p) \tag{P_k}$$
$$\text{s.t. } x \in X, 0 \leq \mathbf{y} \perp \mathbf{M}\mathbf{y} + \mathbf{q}(x) \geq 0,$$

where $F(x, \mathbf{y}) = (f(x, y(\xi^1), \xi^1), f(x, y(\xi^2), \xi^2), \cdots, f(x, y(\xi^k), \xi^k))$,

$$\mathbf{y} = \begin{pmatrix} y(\xi^1) \\ y(\xi^2) \\ \vdots \\ y(\xi^k) \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} M(\xi^1) & 0 & \cdots & 0 \\ 0 & M(\xi^2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M(\xi^k) \end{pmatrix} \text{ and } \mathbf{q}(x) = \begin{pmatrix} q(x, \xi^1) \\ q(x, \xi^2) \\ \vdots \\ q(x, \xi^k) \end{pmatrix}.$$

Moreover, to develop numerical methods and convergence analysis, we consider the following regularized problems of (P) and (P$_k$), respectively,

$$\min_{x,y} \Phi_\epsilon(x, y) := \theta(x) + \sup_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, y(\xi), \xi)])$$
$$\text{s.t. } x \in X, \ 0 \le y(\xi) \perp (M(\xi) + \epsilon I)y(\xi) + q(x, \xi) \ge 0 \text{ for a.e. } \xi \in \varXi \tag{P$_\epsilon$}$$

and

$$\min_{x,\mathbf{y}} \Phi_{\epsilon,k}(x, \mathbf{y}) := \theta(x) + \max_{p \in \mathcal{P}_k} h(F(x, \mathbf{y})p)$$
$$\text{s.t. } x \in X, \ 0 \le \mathbf{y} \perp (\mathbf{M} + \epsilon I)\mathbf{y} + \mathbf{q}(x) \ge 0, \tag{P$_{\epsilon,k}$}$$

where $\epsilon > 0$ is the regularization parameter and I is the identity matrix with proper dimension.

Since $M(\xi)$ is positive semidefinite for fixed ξ, the complementarity problem LCP$(M(\xi) + \epsilon I, q(x, \xi))$ has a unique solution [7], denoted by $\hat{y}_\epsilon(x, \xi)$. Moreover, from the Lipschitz continuity of $q(\cdot, \xi)$, $\hat{y}_\epsilon(\cdot, \xi)$ is Lipschitz continuous [5]. Analogously, LCP$(\mathbf{M} + \epsilon I, \mathbf{q}(x))$ has also a unique solution, denoted by $\hat{\mathbf{y}}_\epsilon(x)$, which is also Lipschitz continuous with respect to (w.r.t.) x.

Mathematical programming with equilibrium constraints (MPEC) has been extensively studied [13, 20, 24]. Structural properties, discrete approximation based on sampling and numerical methods of stochastic MPEC with deterministic probability distribution have been investigated [16, 17, 19, 35, 41]. To the best of our knowledge, there is little discussion on distributionally robust MPEC. Moreover, due to the complementarity constraints and the composite structure of the objective function, the minimax problems (P), (P$_\epsilon$) and (P$_{\epsilon,k}$) are generally nonconvex-nonconcave and their saddle points may not exist. We will focus on their minimax points, minimizers in x-space and corresponding optimality conditions.

The main contributions of the paper are summarized as follows.

- Inspired by the constructive reformulations of pure characteristics demand models in [29], we propose a DRMP-SCC model (P) under ambiguity of probability distributions. We give the definitions of global and local minimax points to capture the optima of the nonconvex-nonconcave minimax problem (P). Some sufficient conditions for the existence of solutions are derived.
- Under certain conditions, we prove the convergence of problem (P$_{\epsilon,k}$) to problem (P) regarding optimal solution sets and optimal values as the regularization parameter $\epsilon \downarrow 0$ and the sample size $k \to \infty$.

– We define stationary points of problems (P), (P$_k$) and (P$_{\epsilon,k}$) in the block coordinatewise sense and establish the convergence of stationary points of problem (P$_{\epsilon,k}$) to those of problem (P) as the regularization parameter $\epsilon \downarrow 0$ and the sample size $k \to \infty$.

Notations. $\mathbb{B}$ denotes the closed unit ball centered at original point in the corresponding space. $\mathbb{R}^n_+$ denotes the set of nonnegative vectors in $\mathbb{R}^n$. $\|\cdot\|$ denotes the Euclidean norm of vectors or the induced matrix norm. $\mathrm{diam}(X) := \sup_{x,z \in X} \|x - z\|$ denotes the diameter of X. $\mathrm{d}(x, Z) = \inf_{z \in Z} \|x - z\|$ and $\mathrm{d}(X, Z) = \sup_{x \in X} \inf_{z \in Z} \|x - z\|$ for $x, z \in \mathbb{R}^n$ and $X, Z \subseteq \mathbb{R}^n$. $\mathrm{int}(X)$ denotes the interior of X.

This paper is organized as follows. In Section 2, we give some preliminaries on the ambiguity set $\mathcal{P}$ and its approximation $\mathcal{P}_k$. In Section 3, we give definitions of global and local minimax points of problems (P), (P$_\epsilon$) and (P$_{\epsilon,k}$), and some existence results. After that, we prove the convergence of the solution set and optimal value of problem (P$_{\epsilon,k}$) to those of problem (P) as $\epsilon \downarrow 0$ and $k \to \infty$. In Section 4, we first give definitions of stationary points of problems (P), (P$_k$) and (P$_{\epsilon,k}$) and then study convergence assertions on the stationary points of problem (P$_{\epsilon,k}$) to those of problem (P). In Section 5, we report numerical results on the pure characteristics demand model under uncertainties of probability distributions. In Section 6, we give concluding remarks.

2 Preliminaries

Note that for each $p \in \mathcal{P}_k$, it uniquely determines a discrete probability distribution $\sum_{i=1}^k p_i \mathbf{1}_{\xi^i}$, where $\mathbf{1}_{\xi^i}(\cdot)$ is the indicator function, namely, $\mathbf{1}_{\xi^i}(\xi) = 1$ if $\xi = \xi^i$; 0 otherwise. Thus, in what follows, we will not distinguish p and the corresponding probability distribution. In other words, we can write

$$\mathcal{P}_k = \left\{ P \in \mathcal{P}(\Xi^k) : \mathbb{E}_P[\Psi(\xi)] \in \Gamma \right\}.$$

Based on the set $\Xi^k = \{\xi^1, \cdots, \xi^k\}$, we have the corresponding Voronoi tessellation of Ξ,

$$\Xi_i := \left\{ \xi \in \Xi : \left\| \xi - \xi^i \right\| = \min_{1 \le j \le k} \left\| \xi - \xi^j \right\| \right\}, \ i = 1, \cdots, k.$$

Obviously, $\Xi = \bigcup_{i=1}^k \Xi_i$ and $\mathrm{int}(\Xi_i) \cap \mathrm{int}(\Xi_j) = \emptyset$ for any $i \ne j$.

In this paper, we make a commonly employed Slater type assumption for ambiguity set (5) as follows (see e.g. [18,40]).

Assumption 1 *There exist $P_0 \in \mathcal{P}(\Xi)$ and $\alpha > 0$ such that*

$$\mathbb{E}_{P_0}[\Psi(\xi)] + \alpha \mathbb{B} \subseteq \Gamma.$$

Proposition 1 *Under Assumption 1, there exists $\bar{k} > 0$ such that $\mathcal{P}_k$ is nonempty for any $k \ge \bar{k}$.*

The proof of Proposition 1 mainly follows from [18, Corollary 11]. For completeness, we put the proof in Appendix.

To measure the distance between two probability distributions (and thus two ambiguity sets), we use the well-known Wasserstein metric.

Definition 1 ([15]) Let $\mathcal{H} := \{\hbar : \Xi \to \mathbb{R} : |\hbar(\xi_1) - \hbar(\xi_2)| \leq \|\xi_1 - \xi_2\|\}$. The Wasserstein metric between $P, Q \in \mathcal{P}(\Xi)$ is defined as

$$\mathbb{D}_W(P, Q) = \sup_{\hbar \in \mathcal{H}} |\mathbb{E}_P[\hbar(\xi)] - \mathbb{E}_Q[\hbar(\xi)]|.$$

Notice that Definition 1 gives the definition of Wasserstein metric by using the Kantorovich-Rubinstein theorem. Another definition of Wasserstein metric is based on the joint probability distributions with marginal distributions P and Q, see e.g. [37] for more details.

The deviation distance between $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{P}(\Xi)$ induced by Wasserstein metric is denoted by $\mathbb{D}_W(\mathcal{P}, \mathcal{Q})$, namely,

$$\mathbb{D}_W(\mathcal{P}, \mathcal{Q}) := \sup_{P \in \mathcal{P}} \inf_{Q \in \mathcal{Q}} \mathbb{D}_W(P, Q).$$

Obviously, $\mathcal{P}_k \subseteq \mathcal{P}$, which implies that the deviation distance between $\mathcal{P}_k$ and $\mathcal{P}$ is always zero, i.e. $\mathbb{D}_W(\mathcal{P}_k, \mathcal{P}) = 0$. To derive the estimation of $\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)$, we assume in the rest of this section that Ξ is bounded and denote

$$\Delta := \sup_{P \in \mathcal{P}(\Xi)} \mathbb{D}_W(P, P_0) \leq \text{diam}(\Xi) < +\infty, \tag{6}$$

where P_0 is defined in Assumption 1.

The following Hoffman's type lemma is from [18, Theorem 2].

Lemma 1 ([18]) *Let Assumption 1 hold. Then, for any $Q \in \mathcal{P}(\Xi)$,*

$$\mathbb{D}_W(Q, \mathcal{P}) \leq \frac{\Delta}{\alpha} \mathrm{d}(\mathbb{E}_Q[\Psi(\xi)], \Gamma),$$

where α is defined in Assumption 1 and Δ is defined in (6).

For a sample set $\Xi^k = \{\xi^1, \cdots, \xi^k\}$ and the resulting Voronoi tessellation $\Xi_1, \cdots, \Xi_k$, we call the following probability distribution

$$P_k := \sum_{i=1}^{k} P(\Xi_i)\mathbf{1}_{\xi^i}$$

the Voronoi projection of P. Denote by

$$\beta_k = \max_{\xi \in \Xi} \min_{1 \leq i \leq k} \|\xi - \xi^i\| \tag{7}$$

the Hausdorff distance between Ξ and Ξ^k.

The following lemma gives an estimation of Wasserstein distance between P and its Voronoi projection.

Lemma 2 ([30, Lemma 4.9]) *Let $P \in \mathcal{P}(\Xi)$ and P_k be the Voronoi projection of P. Then $\mathbb{D}_W(P, P_k) \leq \beta_k$.*

Then, we have an estimation of $\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)$ from [18].

Proposition 2 ([18, Theorem 12]) *Suppose that: (i) Assumption 1 holds; (ii) Ψ is Lipschitz continuous; (iii) the sample set Ξ^k satisfies $\beta_k \to 0$ as $k \to \infty$. Then, for sufficiently large k, we have*

$$\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k) \leq L\beta_k$$

for some $L > 0$.

It is noteworthy that [18, Theorem 12] uses Hausdorff distance to quantify the discrepancy of two ambiguity sets. Since $\mathcal{P}_k \subseteq \mathcal{P}$, we have $\mathbb{D}_W(\mathcal{P}_k, \mathcal{P}) = 0$. Therefore, $\mathbb{H}_W(\mathcal{P}_k, \mathcal{P}) := \max\{\mathbb{D}_W(\mathcal{P}_k, \mathcal{P}), \mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)\} = \mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)$.

In the remaining paper, we tacitly assume that both $\mathcal{P}$ and $\mathcal{P}_k$ are nonempty.

3 Optimal values and solutions of problems (P) and ($\mathrm{P}_{\epsilon,k}$)

Problem (P) can be rewritten as

$$\min_{(x,y) \in \mathcal{D}} \max_{P \in \mathcal{P}} g(x, y, P) := \theta(x) + h(\mathbb{E}_P[f(x, y(\xi), \xi)]),$$

where $\mathcal{D} = \{(x, y) : x \in X, y(\xi) \in \mathrm{SOL}(M(\xi), q(x, \xi)), \xi \in \Xi\}$.

Note that the function $g(\cdot, \cdot, P)$ is not convex for a fixed $P \in \mathcal{P}$, and $g(x, y, \cdot)$ is not concave for a fixed tuple $(x, y) \in \mathcal{D}$. This usually implies that

$$\min_{(x,y) \in \mathcal{D}} \max_{P \in \mathcal{P}} g(x, y, P) \neq \max_{P \in \mathcal{P}} \min_{(x,y) \in \mathcal{D}} g(x, y, P).$$

By the saddle point existence theorem [8], g does not have a saddle point $(x^*, y^*, P^*) \in \mathcal{D} \times \mathcal{P}$ that satisfies

$$g(x^*, y^*, P) \leq g(x^*, y^*, P^*) \leq g(x, y, P^*)$$

for any $(x, y, P) \in \mathcal{D} \times \mathcal{P}$. Hence, we define global and local minimax points of problem (P) by using the idea in [14].

Definition 2 We call $(x^*, y^*, P^*) \in \mathcal{D} \times \mathcal{P}$ a global minimax point of problem (P), if it satisfies

$$g(x^*, y^*, P) \leq g(x^*, y^*, P^*) \leq \max_{Q \in \mathcal{P}} g(x, y, Q)$$

for any $(x, y, P) \in \mathcal{D} \times \mathcal{P}$.

We call $(x^*, y^*, P^*) \in \mathcal{D} \times \mathcal{P}$ a local minimax point of problem (P), if there exist $\delta_0 > 0$ and $\varsigma : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\varsigma(\delta) \to 0$ as $\delta \to 0$, such that

$$g(x^*, y^*, P) \leq g(x^*, y^*, P^*) \leq \max_{Q \in \mathcal{P}, \mathbb{D}_W(Q, P^*) \leq \varsigma(\delta)} g(x, y, Q),$$

for any $\delta \in (0, \delta_0]$, $(x, y) \in \mathcal{D}$ satisfying $\|y - y^*\|_{\mathcal{L}_2} := \left(\int_{\Xi} \|y(\xi) - y^*(\xi)\|^2 \,\mathrm{d}\xi\right)^{\frac{1}{2}} \leq \delta$ and $\|x - x^*\| \leq \delta$, and $P \in \mathcal{P}$ satisfying $\mathbb{D}_W(P, P^*) \leq \delta$.

Similarly, we can define global minimax points and local minimax points of problems (P_ϵ), (P_k) and $(P_{\epsilon,k})$, respectively.

Using the notation $\mathcal{D}$, problem (P) can be written as

$$v^* := \min_{(x,y)\in\mathcal{D}} \Phi(x,y). \tag{8}$$

By substituting the unique solution of $\mathrm{LCP}(M(\xi) + \epsilon I, q(x,\xi))$ (denoted by $\hat{y}_\epsilon(x,\xi)$) into the objective function, problem (P_ϵ) can be written as

$$v_\epsilon^* := \min_{x\in X} \Phi_\epsilon(x, \hat{y}_\epsilon(x, \cdot)). \tag{9}$$

Then its discrete approximation problem $(P_{\epsilon,k})$ can be written as

$$v_{\epsilon,k}^* := \min_{x\in X} \Phi_{\epsilon,k}(x, \hat{\mathbf{y}}_\epsilon(x)), \tag{10}$$

where $\hat{\mathbf{y}}_\epsilon(x) := (\hat{y}_\epsilon(x,\xi^1)^\top, \cdots, \hat{y}_\epsilon(x,\xi^k)^\top)^\top$.

Denote $\mathcal{D}^*$, X_ϵ^* and $X_{\epsilon,k}^*$ the optimal solution sets of problems (8), (9) and (10), respectively. Let $X^* = \mathrm{Proj}_x \mathcal{D}^*$ be the projection of $\mathcal{D}^*$ onto x-space. In this section, we provide conditions for the existence of minimizers of problems (P), (P_ϵ) and $(P_{\epsilon,k})$, and prove the convergence of optimal values and optimal solution sets of problem $(P_{\epsilon,k})$ to those of problem (P) as $\epsilon \downarrow 0$ and $k \to \infty$. The convergence analysis is divided into two parts: the convergence of (P_ϵ) to (P) as $\epsilon \downarrow 0$ and the convergence of $(P_{\epsilon,k})$ to (P_ϵ) as $k \to \infty$ for a fixed $\epsilon > 0$.

3.1 Existence of solutions

In this subsection, we provide some sufficient conditions for the existence of global minimax points of problems (P), (P_ϵ) and $(P_{\epsilon,k})$, respectively.

Let $L_f(\xi^i)$ be the Lipschitz modulus of $f(\cdot, \cdot, \xi^i)$ and $L_\epsilon(\xi^i)$ be the Lipschitz modulus of $\hat{y}_\epsilon(\cdot, \xi^i)$. For a fixed $\epsilon > 0$ and $k > 0$, we have

$$\left| \max_{p\in\mathcal{P}_k} h(F(x, \hat{\mathbf{y}}_\epsilon(x))p) - \max_{p\in\mathcal{P}_k} h(F(x', \hat{\mathbf{y}}_\epsilon(x'))p) \right|$$

$$\leq \max_{p\in\mathcal{P}_k} |h(F(x, \hat{\mathbf{y}}_\epsilon(x))p) - h(F(x', \hat{\mathbf{y}}_\epsilon(x'))p)|$$

$$\leq L_h \max_{p\in\mathcal{P}_k} \left\| \sum_{i=1}^k p_i f(x, \hat{y}_\epsilon(x,\xi^i), \xi^i) - \sum_{i=1}^k p_i f(x', \hat{y}_\epsilon(x',\xi^i), \xi^i) \right\|$$

$$\leq L_h \max_{p\in\mathcal{P}_k} \sum_{i=1}^k p_i \left\| f(x, \hat{y}_\epsilon(x,\xi^i), \xi^i) - f(x', \hat{y}_\epsilon(x',\xi^i), \xi^i) \right\|$$

$$\leq L_h \max_{p\in\mathcal{P}_k} \sum_{i=1}^k p_i L_f(\xi^i)(\|x - x'\| + L_\epsilon(\xi^i) \|x - x'\|)$$

$$\leq L_h \max_{1\leq i\leq k} L_f(\xi^i) \left(1 + \max_{1\leq i\leq k} L_\epsilon(\xi^i) \right) \|x - x'\|$$

for any $x, x' \in X$, where the third inequality follows from Jensen's inequality. Therefore, $\Phi_{\epsilon,k}(x, \hat{\mathbf{y}}_\epsilon(x))$ is Lipschitz continuous over the compact and convex set X, and it has a minimizer $x^* \in X$. Moreover, $\mathcal{P}_k$ is closed and bounded due to the continuity of Ψ and $h(F(x^*, \hat{\mathbf{y}}_\epsilon(x^*))p)$ is continuous w.r.t. p. Hence, there is a maximizer p^* such that (x^*, p^*) is a global minimax point of problem $(\mathrm{P}_{\epsilon,k})$.

In what follows, we provide sufficient conditions for the existence of solutions to problems (P) and (P_ϵ).

Recall that a sequentially compact set means that any sequence contained in this set has a convergent subsequence. We say that a sequence of probability distribution $\{P_j\} \subseteq \mathcal{P}(\Xi)$ weakly converges to $P \in \mathcal{P}(\Xi)$ if $\mathbb{E}_{P_j}[g(\xi)] \to \mathbb{E}_P[g(\xi)]$ as $j \to \infty$ for any continuous and bounded function $g : \Xi \to \mathbb{R}$. A subset of $\mathcal{P}(\Xi)$ is weakly compact if any sequence in this subset has a weakly convergent subsequence with a weak limit point contained in this subset.

Assumption 2

(i) *There exists $\kappa : \Xi \to \mathbb{R}_+$ with $\max_{P \in \mathcal{P}} \mathbb{E}_P[\kappa(\xi)^2] < \infty$ such that*

$$\max_{(x,y) \in \mathcal{D}} \|y(\xi)\| < \kappa(\xi), \quad \xi \in \Xi.$$

(ii) *There exists a tuple $(x_0, y_0) \in \mathcal{D}$ such that $\Phi(x_0, y_0) < \infty$.*

The following theorem provides sufficient conditions for the existence of minimizers of problem (P).

Theorem 1 *Under Assumption 2, if $\mathcal{D}$ is a sequentially compact set, then the objective function of problem (P) satisfies $|\Phi(x, y)| < \infty$ for any $(x, y) \in \mathcal{D}$ and problem (P) has a minimizer (x^*, y^*).*

Proof By Assumption 2, we have

$$
\begin{aligned}
&|\Phi(x, y) - \Phi(x_0, y_0)| \\
&\leq |\theta(x) - \theta(x_0)| + L_h \max_{P \in \mathcal{P}} |\mathbb{E}_P[\|f(x, y(\xi), \xi)\|] - \mathbb{E}_P[\|f(x_0, y_0(\xi), \xi)\|]| \\
&\leq L_\theta \mathrm{diam}(X) + L_h \max_{P \in \mathcal{P}} \mathbb{E}_P[L_f(\xi)(\mathrm{diam}(X) + 2\kappa(\xi))] < \infty.
\end{aligned}
$$

Hence $|\Phi(x, y)| < \infty$ for any $(x, y) \in \mathcal{D}$, which means that there is $v^* \in \mathbb{R}$ such that $v^* = \inf_{(x,y) \in \mathcal{D}} \Phi(x, y)$. To prove that problem (P) has a minimizer (x^*, y^*) under Assumption 2, we only need to show that Φ is lower semicontinuous (lsc) due to the sequential compactness of $\mathcal{D}$ [32, Theorem 1.9].

First, we show that for a fixed $x \in X$, $\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \cdot, \xi)])$ is lsc. By the Lipschitz continuity of h and f, we know that

$$
\begin{aligned}
&|h(\mathbb{E}_P[f(x, y'(\xi), \xi)]) - h(\mathbb{E}_P[f(x, y(\xi), \xi)])| \\
&\leq L_h \mathbb{E}_P[L_f(\xi)\|y'(\xi) - y(\xi)\|] \\
&\leq L_h (\mathbb{E}_P[L_f^2(\xi)])^{\frac{1}{2}} \left(\mathbb{E}_P[\|y'(\xi) - y(\xi)\|^2]\right)^{\frac{1}{2}} \to 0
\end{aligned}
$$

if $\left(\mathbb{E}_P[\|y'(\xi) - y(\xi)\|^2]\right)^{\frac{1}{2}} \to 0$. Hence $h(\mathbb{E}_P[f(x,\cdot,\xi)])$ is continuous for fixed $x \in X$ and $P \in \mathcal{P}$, which implies that $\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x,\cdot,\xi)])$ is lsc.

Similarly, we can show that $\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(\cdot,\cdot,\xi)])$ is lsc over $\mathcal{D}$. Since θ is a Lipchitz continuous function of x, we derive that Φ is lsc. $\qquad\square$

The following theorem establishes the existence of minimizers and global minimax points of problem (P_ϵ).

Theorem 2 *Suppose that Ξ is bounded. Then for any $\epsilon > 0$, problem (P_ϵ) has a minimizer $(x^*, \hat{y}_\epsilon(x^*, \cdot))$. In addition, if $\mathcal{P}$ is weakly compact, then problem (P_ϵ) has a global minimax point $(x^*, \hat{y}_\epsilon(x^*, \cdot), P^*)$.*

Proof For the existence of a minimizer $(x^*, \hat{y}_\epsilon(x^*, \cdot))$, we only need to show Assumption 2 holds with Φ_ϵ and

$$\mathcal{D}_\epsilon = \{(x, y) : x \in X, \ y(\xi) \in \mathrm{SOL}(M(\xi) + \epsilon I, q(x, \xi)), \xi \in \Xi\}$$

for any $\epsilon > 0$.

Since $M(\xi) + \epsilon I$ is a positive definite matrix for any $\xi \in \Xi$, the solution set $\mathrm{SOL}(M(\xi) + \epsilon I, q(x, \xi))$ has a unique vector $\hat{y}_\epsilon(x, \xi)$ for any $x \in X$ and $\xi \in \Xi$. Moreover, from Lipschitz continuity of q and M and boundedness of X and Ξ, there is a positive number $\Lambda_\epsilon > 0$ such that

$$\Lambda_\epsilon \geq \max_{x \in X, \xi \in \Xi} \max_{d \in [0,1]^n} \|(I - D + D(M(\xi) + \epsilon I))^{-1}\| \|\min(0, q(x, \xi))\|,$$

which implies $\|\hat{y}_\epsilon(x, \xi)\| \leq \Lambda_\epsilon$, for $x \in X, \xi \in \Xi$. See error bounds for the LCP in [5,7]. Hence, Assumption 2 holds with Φ_ϵ and $\mathcal{D}_\epsilon$. Using a similar proof of Theorem 1, we can show that problem (P_ϵ) has a minimizer $(x^*, \hat{y}_\epsilon(x^*, \cdot))$.

Moreover, if $\mathcal{P}$ is weakly compact, due to the continuity and boundedness of $f(x^*, \hat{y}_\epsilon(x^*, \xi), \xi)$ w.r.t. ξ, the following maximization problem

$$\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x^*, \hat{y}_\epsilon(x^*, \xi), \xi)])$$

has a maximizer P^*. Hence $(x^*, \hat{y}_\epsilon(x^*, \cdot), P^*)$ is a global minimax point of problem (P_ϵ). $\qquad\square$

3.2 Convergence analysis between problems (P) and (P_ϵ)

We use that $\mathrm{LCP}(M(\xi) + \epsilon I, q(x, \xi))$ has a unique solution $\hat{y}_\epsilon(x, \xi)$ to introduce the following auxiliary problem with fixed $\epsilon > 0$ and $r_\epsilon > 0$,

$$\min_{x, y} \Phi(x, y) \quad \text{s.t. } x \in X, \ y \in \mathcal{M}_\epsilon(x) := \left\{ y : \sup_{P \in \mathcal{P}} \mathbb{E}_P[\|\hat{y}_\epsilon(x, \xi) - y(\xi)\|^2] \leq r_\epsilon \right\}. \tag{11}$$

Since $\hat{y}_\epsilon(x, \cdot)$ is a measurable function for any $x \in X$ and X is bounded, $y \in \mathcal{M}_\epsilon(x)$ is a measurable function for any $x \in X$. By assumptions of Theorems 1 and 2, problems (P), (P_ϵ) and (11) have minimizers. Denote by ϑ_ϵ the optimal value of problem (11). We make the following technical assumption.

Assumption 3 $|\vartheta_\epsilon - v^*| \to 0$ *as* $\epsilon \downarrow 0$.

Remark 1 Assumption 3 is a standard assumption and has been used in the perturbation analysis for parametric programming [3, Chapter 4]. In [4], a stronger assumption is used for problem (1) (cf. [4, Assumption 1]), that is, for any $\delta > 0$ there exists an ϵ_0 such that for any $\epsilon \in [0, \epsilon_0]$,

$$D^\epsilon \cap (X^* + \mathbb{B}(0, \delta)) \neq \emptyset \tag{12}$$

holds, where $D^\epsilon := \{x \in X : \|A_t \mathbb{E}_P[s_{t,\epsilon}(x, \xi)] - b_t\| \leq r(\epsilon), t = 1, \cdots, T\}$, $s_{t,\epsilon}(x, \xi)$ is the unique regularization solution of problem (1) for fixed t, x and ξ, $r(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$, X^* is the optimal solution set of problem (1). Assumption on the optimal solution set of perturbation problem (11) of problem (P) in the way (12) implies Assumption 3 on their optimal values. However, the converse is not true.

The following theorem provides the relationship between optimal values of problems (P) and (P$_\epsilon$) as $\epsilon \downarrow 0$.

Theorem 3 *Under Assumption 3, if $r_\epsilon \to 0$ as $\epsilon \downarrow 0$, then we have*

$$|v_\epsilon^* - v^*| \to 0 \ as \ \epsilon \downarrow 0.$$

Proof Using the Lipschitz continuity conditions on $\theta, h, f(\cdot, \cdot, \xi)$, we have

$$\begin{aligned}
|v_\epsilon^* - \vartheta_\epsilon| &= \left| \min_{x \in X} \left(\theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right) \right. \\
&\quad \left. - \min_{x \in X} \left(\theta(x) + \min_{y \in \mathcal{M}_\epsilon(x)} \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, y(\xi), \xi)]) \right) \right| \\
&\leq \max_{x \in X} \left(\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \min_{y \in \mathcal{M}_\epsilon(x)} \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, y(\xi), \xi)]) \right) \\
&= \max_{x \in X} \max_{y \in \mathcal{M}_\epsilon(x)} \left(\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, y(\xi), \xi)]) \right) \\
&\leq L_h \sup_{x \in X, y \in \mathcal{M}_\epsilon(x), P \in \mathcal{P}} \mathbb{E}_P[L_f(\xi) \|\hat{y}_\epsilon(x, \xi) - y(\xi)\|],
\end{aligned}$$

where the first inequality follows from the fact that

$$\hat{y}_\epsilon(x, \xi) \in \mathcal{M}_\epsilon(x) := \left\{ y : \sup_{P \in \mathcal{P}} \mathbb{E}_P[\|\hat{y}_\epsilon(x, \xi) - y(\xi)\|^2] \leq r_\epsilon \right\}.$$

By Hölder's inequality, we have that

$$\sup_{x \in X, y \in \mathcal{M}_\epsilon(x), P \in \mathcal{P}} \mathbb{E}_P[L_f(\xi) \|\hat{y}_\epsilon(x, \xi) - y(\xi)\|] \leq \sqrt{r_\epsilon} \left(\sup_{P \in \mathcal{P}} \mathbb{E}_P[L_f^2(\xi)] \right)^{\frac{1}{2}} \to 0$$

as $\epsilon \downarrow 0$ due to $\sup_{P \in \mathcal{P}} \mathbb{E}_P[L_f^2(\xi)] < \infty$ and $r_\epsilon \to 0$ as $\epsilon \downarrow 0$. Thus, we obtain $|v_\epsilon^* - \vartheta_\epsilon| \to 0$. This together with Assumption 3 implies that $|v_\epsilon^* - v^*| \to 0$ as $\epsilon \downarrow 0$. $\qquad\square$

Since $M(\xi)$ is positive semidefinite and the solution set of $\mathrm{LCP}(M(\xi), q(x, \xi))$ is nonempty for any $x \in X$ and $\xi \in \Xi$ by the assumption of relatively complete recourse, the unique solution $\hat{y}_\epsilon(x, \xi)$ converges to the least norm solution $\bar{y}(x, \xi)$ of $\mathrm{LCP}(M(\xi), q(x, \xi))$ as $\epsilon \downarrow 0$ [7, Theorem 5.6.2], which is defined as

$$\bar{y}(x, \xi) = \arg\min \ \{\|y(x, \xi)\| : y(x, \xi) \in \mathrm{SOL}(M(\xi), q(x, \xi))\}.$$

The following proposition gives a sufficient condition for the convergence of optimal values of problem (P_ϵ) to that of problem (P) as $\epsilon \downarrow 0$.

Proposition 3 *Suppose that* $\sup_{P \in \mathcal{P}} \mathbb{E}_P[\|\hat{y}_\epsilon(x, \xi) - \bar{y}(x, \xi)\|^2] \to 0$ *uniformly as* $\epsilon \downarrow 0$ *in* X *and* $v^* = \min_{x \in X} (\theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \bar{y}(x, \xi), \xi)]))$. *Then*

$$|v_\epsilon^* - v^*| \to 0 \ \text{and} \ \mathrm{d}(X_\epsilon^*, X^*) \to 0 \ \text{as} \ \epsilon \downarrow 0.$$

Proof By assumptions and the boundedness of X, we have

$$
\begin{aligned}
|v_\epsilon^* - v^*| &= \left| \min_{x \in X} \left(\theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right) \right. \\
&\quad \left. - \min_{x \in X} \left(\theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \bar{y}(x, \xi), \xi)]) \right) \right| \\
&\leq \max_{x \in X} \left| \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \bar{y}(x, \xi), \xi)]) \right| \\
&\leq L_h \max_{x \in X} \max_{P \in \mathcal{P}} \mathbb{E}_P[\|f(x, \hat{y}_\epsilon(x, \xi), \xi)] - \mathbb{E}_P[f(x, \bar{y}(x, \xi), \xi)\|] \\
&\leq L_h \max_{x \in X} \max_{P \in \mathcal{P}} \mathbb{E}_P[L_f(\xi) \|\hat{y}_\epsilon(x, \xi) - \bar{y}(x, \xi)\|] \\
&\leq L_h \max_{x \in X} \left(\sup_{P \in \mathcal{P}} \mathbb{E}_P[L_f^2(\xi)] \right)^{\frac{1}{2}} \left(\sup_{P \in \mathcal{P}} \mathbb{E}_P[\|\hat{y}_\epsilon(x, \xi) - \bar{y}(x, \xi)\|^2] \right)^{\frac{1}{2}} \to 0
\end{aligned}
$$

as $\epsilon \downarrow 0$, where the second inequality follows from Jensen's inequality and the last inequality follows from Hölder's inequality. Moreover, from the above inequalities, we find that $\Phi_\epsilon(x, \hat{y}_\epsilon(x, \cdot))$ converges to $\Phi(x, \bar{y}(x, \cdot))$ uniformly w.r.t. x over X as $\epsilon \downarrow 0$. By [34, Theorem 5.3], we derive the convergence of $\mathrm{d}(X_\epsilon^*, X^*)$. $\qquad\square$

3.3 Convergence analysis between problems (P_ϵ) and $(\mathrm{P}_{\epsilon,k})$

In this subsection, for a fixed $\epsilon > 0$, we consider the convergence between problems (P_ϵ) and $(\mathrm{P}_{\epsilon,k})$ as $k \to \infty$. Let the support set Ξ be bounded throughout this subsection.

By [5, Theorem 2.8], there exists an $\alpha > 0$ such that for any $\xi_1, \xi_2 \in \Xi$,

$$\max_{x \in X} \|\hat{y}_\epsilon(x, \xi_1) - \hat{y}_\epsilon(x, \xi_2)\| \leq \frac{\alpha}{\epsilon} \max_{x \in X, \xi \in \Xi} \|q(x, \xi)\| \|\xi_1 - \xi_2\| =: \hat{L}\|\xi_1 - \xi_2\|,$$

where the existence of a constant $\hat{L} > 0$ employs the boundedness of X and Ξ and Lipschitz continuity of q. Moreover, if f is Lipschitz continuous, there exists an $\bar{L}_f > 0$ such that for any $x \in X$,

$$\|f(x, \hat{y}_\epsilon(x, \xi_1), \xi_1) - f(x, \hat{y}_\epsilon(x, \xi_2), \xi_2)\| \leq \bar{L}_f \|\xi_1 - \xi_2\|. \tag{13}$$

We give the following quantitative stability results between problems (9) and (10) based on Wasserstein metric.

Theorem 4 *Let f be Lipschitz continuous. There exists an $L > 0$ such that*

$$\left|v_{\epsilon,k}^* - v_\epsilon^*\right| \leq L\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k), \tag{14}$$

$$d(X_{\epsilon,k}^*, X_\epsilon^*) \leq \mathcal{R}^{-1}(2L\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)), \tag{15}$$

where $\mathcal{R} : \mathbb{R}_+ \to \mathbb{R}_+$ is the growth function of problem (P_ϵ), i.e.,

$$\mathcal{R}(\tau) := \min \left\{ \theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - v_\epsilon^* : d(x, X_\epsilon^*) \geq \tau, x \in X \right\}$$

and its inverse

$$\mathcal{R}^{-1}(t) := \sup\{\tau \in \mathbb{R}_+ : \mathcal{R}(\tau) \leq t\}.$$

Proof Using (13), we have

$$\left|v_\epsilon^* - v_{\epsilon,k}^*\right|$$

$$= \left| \min_{x \in X} \left(\theta(x) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right) \right.$$

$$\left. - \min_{x \in X} \left(\theta(x) + \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right) \right|$$

$$\leq \max_{x \in X} \left| \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right|$$

$$= \max_{x \in X} \left(\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right) \tag{16}$$

$$= \max_{x \in X} \max_{P \in \mathcal{P}} \min_{Q \in \mathcal{P}_k} (h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - h(\mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]))$$

$$\leq \max_{x \in X} L_h \max_{P \in \mathcal{P}} \min_{Q \in \mathcal{P}_k} \|\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)] - \mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]\|$$

$$\leq L_h \bar{L}_f \mathbb{D}_W(\mathcal{P}, \mathcal{P}_k), \tag{17}$$

where the second equality follows from $\mathcal{P}_k \subseteq \mathcal{P}$, and the last inequality follows from Definition 1. We obtain (14) by setting $L := L_h \bar{L}_f$.

From (16) and (17), for any $\tilde{x} \in X_{\epsilon,k}^*$, we have

$$L\mathbb{D}_W(\mathcal{P}, \mathcal{P}_k)$$

$$\geq \max_{x \in X} \left(\max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) - \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(x, \hat{y}_\epsilon(x, \xi), \xi)]) \right)$$

$$\geq \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(\tilde{x}, \hat{y}_\epsilon(\tilde{x}, \xi), \xi)]) - \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(\tilde{x}, \hat{y}_\epsilon(\tilde{x}, \xi), \xi)])$$

$$= \theta(\tilde{x}) + \max_{P \in \mathcal{P}} h(\mathbb{E}_P[f(\tilde{x}, \hat{y}_\epsilon(\tilde{x}, \xi), \xi)]) - \left(\theta(\tilde{x}) + \max_{Q \in \mathcal{P}_k} h(\mathbb{E}_Q[f(\tilde{x}, \hat{y}_\epsilon(\tilde{x}, \xi), \xi)]) \right).$$

By using the above estimation, we obtain

$$
\begin{aligned}
2L\mathbb{D}_W(\mathcal{P},\mathcal{P}_k) &\geq L\mathbb{D}_W(\mathcal{P},\mathcal{P}_k) + \left|v_\epsilon^* - v_{\epsilon,k}^*\right| \\
&\geq L\mathbb{D}_W(\mathcal{P},\mathcal{P}_k) + v_{\epsilon,k}^* - v_\epsilon^* \\
&\geq \theta(\tilde{x}) + \max_{P\in\mathcal{P}} h(\mathbb{E}_P[f(\tilde{x},\hat{y}_\epsilon(\tilde{x},\xi),\xi)]) \\
&\quad - \left(\theta(\tilde{x}) + \max_{Q\in\mathcal{P}_k} h(\mathbb{E}_Q[f(\tilde{x},\hat{y}_\epsilon(\tilde{x},\xi),\xi)])\right) + v_{\epsilon,k}^* - v_\epsilon^* \\
&= \theta(\tilde{x}) + \sup_{P\in\mathcal{P}} h(\mathbb{E}_P[f(\tilde{x},\hat{y}_\epsilon(\tilde{x},\xi),\xi)]) - v_\epsilon^* \\
&\geq \mathcal{R}(\mathrm{d}(\tilde{x},X_\epsilon^*)),
\end{aligned}
$$

which implies that

$$
\mathrm{d}(\tilde{x},X_\epsilon^*) \leq \mathcal{R}^{-1}\left(2L\mathbb{D}_W(\mathcal{P},\mathcal{P}_k)\right).
$$

Due to the arbitrariness of $\tilde{x}\in X_{\epsilon,k}^*$, we derive (15). $\qquad\square$

Since $\mathbb{H}_W(\mathcal{P}_k,\mathcal{P}) := \max\{\mathbb{D}_W(\mathcal{P}_k,\mathcal{P}),\mathbb{D}_W(\mathcal{P},\mathcal{P}_k)\} = \mathbb{D}_W(\mathcal{P},\mathcal{P}_k)$ due to $\mathcal{P}_k \subseteq \mathcal{P}$, we use $\mathbb{D}_W(\mathcal{P},\mathcal{P}_k)$ in Theorem 4 instead of $\mathbb{H}_W(\mathcal{P}_k,\mathcal{P})$.

From the quantitative results in Theorem 4 and Proposition 2, we have the following convergence theorem.

Theorem 5 *Under the conditions of Proposition 2 and Theorem 4, we have*

$$
\left|v_{\epsilon,k}^* - v_\epsilon^*\right| \to 0 \text{ and } \mathrm{d}(X_{\epsilon,k}^*, X_\epsilon^*) \to 0 \text{ as } k\to\infty.
$$

Finally, Theorem 3 together with Theorem 5 entails the following convergence results from $(\mathrm{P}_{\epsilon,k})$ to (P).

Theorem 6 *Suppose that the conditions of Theorems 3 and 5 hold. Then*

$$
\lim_{\epsilon\downarrow 0}\lim_{k\to\infty} v_{\epsilon,k}^* = v^*.
$$

If, moreover, the conditions of Proposition 3 hold, then

$$
\lim_{\epsilon\downarrow 0}\lim_{k\to\infty} \mathrm{d}(X_{\epsilon,k}^*, X^*) = 0.
$$

4 Stationarity of problems (P) and $(\mathbf{P}_{\epsilon,k})$

In this section, we consider the convergence of stationary points of problem $(\mathrm{P}_{\epsilon,k})$ to these of problem (P) as $\epsilon\downarrow 0$ and $k\to\infty$. We first consider the stationary points defined by the regular normal cone in [32].

The regular normal cone to a closed set Ω at $\bar{x}\in\Omega$, denoted by $\widehat{\mathcal{N}}_\Omega(\bar{x})$, is

$$
\widehat{\mathcal{N}}_\Omega(\bar{x}) := \{d : d^\top(x-\bar{x}) \leq o(\|x-\bar{x}\|),\ \forall x\in\Omega\},
$$

where $o(\cdot)$ means that $o(a)/a \to 0$ as $a \downarrow 0$. The (limiting) normal cone to a closed set Ω at $\bar{x} \in \Omega$, denoted by $\mathcal{N}_\Omega(\bar{x})$, is

$$\mathcal{N}_\Omega(\bar{x}) := \{d : \exists\, x^k \in \Omega, x^k \to \bar{x},\ d^k \to d \text{ with } d^k \in \widehat{\mathcal{N}}_\Omega(x^k)\}.$$

$\widehat{\mathcal{N}}_\Omega(\bar{x})$ is a closed and convex cone and $\widehat{\mathcal{N}}_\Omega(\cdot)$ is outer semicontinuous over Ω. $\mathcal{N}_\Omega(\bar{x})$ is a closed cone. Generally, we have $\widehat{\mathcal{N}}_\Omega(\bar{x}) \subseteq \mathcal{N}_\Omega(\bar{x})$. They are consistent when Ω is convex, namely,

$$\widehat{\mathcal{N}}_\Omega(\bar{x}) = \mathcal{N}_\Omega(\bar{x}) = \{d : d^\top(x - \bar{x}) \leq 0,\ \forall x \in \Omega\}.$$

A necessary condition for $\bar{x}$ to be a local minimizer of $\min_{x \in \Omega}\ \phi(x)$, where ϕ is a differentiable function over Ω, is (see [32, Theorem 6.12])

$$0 \in \nabla\phi(\bar{x}) + \widehat{\mathcal{N}}_\Omega(\bar{x}).$$

In Section 4.1, we focus on the concepts of stationary points of problems (P), (P$_k$) and (P$_{\epsilon,k}$). In Section 4.2, we give convergence results regarding the stationary points between problems (P$_{\epsilon,k}$) and (P) as $\epsilon \downarrow 0$ and $k \to \infty$.

4.1 Concepts of stationarity

We first consider the concepts of stationary points of the discrete problems (P$_k$) and (P$_{\epsilon,k}$). To simplify the notation, we denote

$$\begin{aligned}
&G(x, \mathbf{y}, p) = \theta(x) + h(F(x, \mathbf{y})p), \\
&\mathcal{Z} = \{(x, \mathbf{y}) : x \in X,\ \mathbf{y} \in \mathrm{SOL}(\mathbf{M}, \mathbf{q}(x))\}, \\
&\mathcal{Z}(x) = \{\mathbf{y} : (x, \mathbf{y}) \in \mathcal{Z}\}, \quad \mathcal{Z}(\mathbf{y}) = \{x : (x, \mathbf{y}) \in \mathcal{Z}\}, \\
&\mathcal{Z}_\epsilon = \{(x, \mathbf{y}) : x \in X,\ \mathbf{y} \in \mathrm{SOL}(\mathbf{M} + \epsilon I, \mathbf{q}(x))\}, \\
&\mathcal{Z}_\epsilon(x) = \{\mathbf{y} : (x, \mathbf{y}) \in \mathcal{Z}_\epsilon\}, \quad \mathcal{Z}_\epsilon(\mathbf{y}) = \{x : (x, \mathbf{y}) \in \mathcal{Z}_\epsilon\}.
\end{aligned}$$

Then problems (P$_k$) and (P$_{\epsilon,k}$) can be rewritten, respectively, as

$$\min_{(x,\mathbf{y}) \in \mathcal{Z}} \max_{p \in \mathcal{P}_k}\ G(x, \mathbf{y}, p) \tag{18}$$

and

$$\min_{(x,\mathbf{y}) \in \mathcal{Z}_\epsilon} \max_{p \in \mathcal{P}_k}\ G(x, \mathbf{y}, p). \tag{19}$$

Note that the set $\mathcal{P}_k$ is convex and bounded. We have the following optimality condition for a local minimax point.

Proposition 4 (optimality condition for a local minimax point) *If $(\bar{x}, \bar{\mathbf{y}}, \bar{p}) \in \mathcal{Z} \times \mathcal{P}_k$ is a local minimax point of problem* (18), *then it satisfies*

$$\begin{cases} 0 \in \nabla_{(x,\mathbf{y})} G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \widehat{\mathcal{N}}_{\mathcal{Z}}(\bar{x}, \bar{\mathbf{y}}), \\ 0 \in -\nabla_p G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{P}_k}(\bar{p}). \end{cases} \tag{20}$$

Proof Let $(\bar{x}, \bar{\mathbf{y}}, \bar{p})$ be a local minimax point. To simplify the notation, denote $\bar{z} = (\bar{x}, \bar{\mathbf{y}})$. According to the definition of local minimax points, there exist a $\delta_0 > 0$ and a function $\varsigma : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\varsigma(\delta) \to 0$ as $\delta \to 0$, such that for any $\delta \in (0, \delta_0]$ and $(z, p) \in \mathcal{Z} \times \mathcal{P}_k$ satisfying $\|z - \bar{z}\| \leq \delta$ and $\|p - \bar{p}\| \leq \delta$, we have

$$G(\bar{z}, p) \leq G(\bar{z}, \bar{p}) \leq \max_{p' \in \{p \in \mathcal{P}_k : \|p - \bar{p}\| \leq \varsigma(\delta)\}} G(z, p'). \tag{21}$$

Obviously, the first inequality in (21) implies the second assertion in (20).

Next, we verify the first assertion in (20). For any $z \in \mathcal{Z}$, let $\mathcal{T}_{\mathcal{Z}}(z)$ be the tangent cone of $\mathcal{Z}$ at z (see [32, Definition 6.1]) and

$$\tilde{p}(z) \in \underset{p' \in \{p \in \mathcal{P}_k : \|p - \bar{p}\| \leq \varsigma(\delta)\}}{\arg \max} G(z, p').$$

From the second inequality in (21), we have, for any $w \in \mathcal{T}_{\mathcal{Z}}(z)$ with $\|w\| = \delta$ and $\bar{z} + w \in \mathcal{Z}$, that

$$\begin{aligned}
0 &\leq G(\bar{z} + w, \tilde{p}(\bar{z} + w)) - G(\bar{z}, \bar{p}) \\
&\leq G(\bar{z} + w, \tilde{p}(\bar{z} + w)) - G(\bar{z}, \tilde{p}(\bar{z} + w)) + G(\bar{z}, \tilde{p}(\bar{z} + w)) - G(\bar{z}, \bar{p}) \\
&\leq G(\bar{z} + w, \tilde{p}(\bar{z} + w)) - G(\bar{z}, \tilde{p}(\bar{z} + w)) \\
&= \nabla_z G(\bar{z}, \tilde{p}(\bar{z} + w))^\top w + o(\|w\|) \\
&= \nabla_z G(\bar{z}, \bar{p})^\top w + (\nabla_z G(\bar{z}, \tilde{p}(\bar{z} + w)) - \nabla_z G(\bar{z}, \bar{p}))^\top w + o(\|w\|) \\
&= \nabla_z G(\bar{z}, \bar{p})^\top w + o(\|w\|),
\end{aligned}$$

where the last equality follows from $\|\tilde{p}(\bar{z} + w) - \bar{p}\| \leq \varsigma(\delta) \to 0$ as $\delta \to 0$, which implies

$$(\nabla_z G(\bar{z}, \tilde{p}(\bar{z} + w)) - \nabla_z G(\bar{z}, \bar{p}))^\top w = o(\|w\|).$$

Thus, $\nabla_z G(\bar{z}, \bar{p})^\top w \geq 0$ for any $w \in \mathcal{T}_{\mathcal{Z}}(z)$, which indicates (see [32, Proposition 6.5]) the first assertion in (20). □

It is noteworthy that the necessary condition for local minimax points in Proposition 4 is also a necessary condition for saddle points, see e.g. [31,28, 27]. $\mathcal{Z}(\bar{x})$ is a convex set because of the positive semidefiniteness of $\mathbf{M}$. If we consider $(x, \mathbf{y})$ individually, it then leads to the concept of block coordinatewise stationarity (see e.g. [42]) as follows:

$$\begin{cases} 0 \in \nabla_x G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \widehat{\mathcal{N}}_{\mathcal{Z}(\bar{\mathbf{y}})}(\bar{x}), \\ 0 \in \nabla_{\mathbf{y}} G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{Z}(\bar{x})}(\bar{\mathbf{y}}), \\ 0 \in -\nabla_p G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{P}_k}(\bar{p}). \end{cases} \tag{22}$$

Condition (22) is a weaker necessary condition than (20) for local optimality of problem (18) (see e.g. [42, Remark 2.2]). The second assertion in (22)

corresponds to a necessary condition for local optimality of the mathematical programming with linear complementarity constraint (MPLCC). Consider

$$\min_{\mathbf{y}} G(\bar{x}, \mathbf{y}, \bar{p})$$
$$\text{s.t. } 0 \leq \mathbf{y} \perp (\mathbf{M} + \epsilon\mathbf{I})\mathbf{y} + \mathbf{q}(\bar{x}) \geq 0. \tag{23}$$

To characterize optimality conditions of MPLCC, we define the following index sets:

$$\mathcal{I}_{+0}(\mathbf{y}) = \{i : \mathbf{y}_i > 0, \ ((\mathbf{M} + \epsilon\mathbf{I})\mathbf{y} + \mathbf{q}(\bar{x}))_i = 0\},$$
$$\mathcal{I}_{0+}(\mathbf{y}) = \{i : \mathbf{y}_i = 0, \ ((\mathbf{M} + \epsilon\mathbf{I})\mathbf{y} + \mathbf{q}(\bar{x}))_i > 0\},$$
$$\mathcal{I}_{00}(\mathbf{y}) = \{i : \mathbf{y}_i = 0, \ ((\mathbf{M} + \epsilon\mathbf{I})\mathbf{y} + \mathbf{q}(\bar{x}))_i = 0\}.$$

Definition 3 (stationary points for MPLCC, [6]) (i) We say that $\mathbf{y}^*$ is a (weak stationary) W-stationary point of problem (23) if there exist multipliers $(\lambda, \mu) \in \mathbb{R}^{mk} \times \mathbb{R}^{mk}$ such that

$$\nabla_{\mathbf{y}} G(\bar{x}, \mathbf{y}^*, \bar{p}) - \lambda - (\mathbf{M}^\top + \epsilon\mathbf{I})\mu = 0,$$
$$\lambda_i = 0 \text{ for } i \in \mathcal{I}_{+0}(\mathbf{y}^*), \ \mu_i = 0 \text{ for } i \in \mathcal{I}_{0+}(\mathbf{y}^*). \tag{24}$$

(ii) We say that $\mathbf{y}^*$ is a (Clarke stationary) C-stationary point of problem (23) if there exist $(\lambda, \mu) \in \mathbb{R}^{mk} \times \mathbb{R}^{mk}$ satisfying (24) and

$$\lambda_i\mu_i \geq 0 \text{ for } i \in \mathcal{I}_{00}(\mathbf{y}^*).$$

(iii) We say that $\mathbf{y}^*$ is an (Mordukhovich stationary) M-stationary point of problem (23) if there exist $(\lambda, \mu) \in \mathbb{R}^{mk} \times \mathbb{R}^{mk}$ satisfying (24) and

$$\lambda_i > 0, \ \mu_i > 0 \text{ or } \lambda_i\mu_i = 0 \text{ for } i \in \mathcal{I}_{00}(\mathbf{y}^*).$$

(iv) We say that $\mathbf{y}^*$ is an (strong stationary) S-stationary point of problem (23) if there exist $(\lambda, \mu) \in \mathbb{R}^{mk} \times \mathbb{R}^{mk}$ satisfying (24) and

$$\lambda_i \geq 0, \ \mu_i \geq 0 \text{ for } i \in \mathcal{I}_{00}(\mathbf{y}^*).$$

Obviously, we have the following observation:

S-stationarity $\Rightarrow$ M-stationarity $\Rightarrow$ C-stationarity $\Rightarrow$ W-stationarity.

For more details about the optimality conditions of the mathematical programming with equilibrium constraints (MPEC), we refer to monograph [20, Chapter 3].

Combining (22) with Definition 3, we give the following concepts of block coordinatewise stationary points of problem (P_k).

Definition 4 (block coordinatewise •-stationary points of (P_k)**)** Let $(\bar{x}, \bar{\mathbf{y}}, \bar{p}) \in \mathcal{Z} \times \mathcal{P}_k$. We say that it is a block coordinatewise •-stationary point of problem (18) if it satisfies

$$
\begin{cases}
0 \in \nabla_x G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \widehat{\mathcal{N}}_{\mathcal{Z}(\bar{\mathbf{y}})}(\bar{x}), \\
\bar{\mathbf{y}} \text{ is a •-stationary point of problem (23) with } \epsilon = 0, \\
0 \in -\nabla_p G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{P}_k}(\bar{p}),
\end{cases}
$$

where "•" can be W, C, M and S.

Analogously, we give the block coordinatewise stationary point of the regularized problem $(P_{\epsilon,k})$. Since there is a unique solution of $\text{LCP}(\mathbf{M} + \epsilon I, \mathbf{q}(\bar{x}))$ for any $\bar{x} \in X$, we use the following definition of block coordinatewise stationary points for problem $(P_{\epsilon,k})$.

Definition 5 (block coordinatewise stationary points of $(P_{\epsilon,k})$**)** We call $(\bar{x}, \bar{\mathbf{y}}, \bar{p}) \in \mathcal{Z}_\epsilon \times \mathcal{P}_k$ a block coordinatewise stationary point of problem (19), if it satisfies

$$
\begin{cases}
0 \in \nabla_x G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \widehat{\mathcal{N}}_{\mathcal{Z}_\epsilon(\bar{\mathbf{y}})}(\bar{x}), \\
\bar{\mathbf{y}} \in \text{SOL}(\mathbf{M} + \epsilon I, \mathbf{q}(\bar{x})), \\
0 \in -\nabla_p G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{P}_k}(\bar{p}).
\end{cases}
\tag{25}
$$

The following lemma shows that a point satisfying (25) must be a block coordinatewise S-stationary point of problem $(P_{\epsilon,k})$.

Lemma 3 *For fixed* $\bar{x} \in X$ *and* $\bar{p} \in \mathcal{P}_k$, $\hat{\mathbf{y}}_\epsilon(\bar{x})$ *is an S-stationary point of the MPLCC* (23).

Proof The feasible set of (23) with $\epsilon > 0$ has a unique vector $\hat{\mathbf{y}}_\epsilon(\bar{x})$. From [11, Proposition 2.2, (ii) and (iii)], $\hat{\mathbf{y}}_\epsilon(\bar{x})$ is an S-stationary point since both $\mathbf{y}$ and $(\mathbf{M} + \epsilon \mathbf{I})\mathbf{y} + q(\bar{x})$ are linear functions of $\mathbf{y}$.   □

Remark 2 Note that $\mathcal{Z}_\epsilon(\bar{\mathbf{y}}) \subseteq X$ and thus $\mathcal{N}_X(\bar{x}) \subseteq \widehat{\mathcal{N}}_{\mathcal{Z}_\epsilon(\bar{\mathbf{y}})}(\bar{x})$. Hence, if $(\bar{x}, \bar{\mathbf{y}}, \bar{p}) \in \mathcal{Z}_\epsilon \times \mathcal{P}_k$ satisfies

$$
\begin{cases}
0 \in \nabla_x G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_X(\bar{x}), \\
\bar{\mathbf{y}} \in \text{SOL}(\mathbf{M} + \epsilon I, \mathbf{q}(\bar{x})), \\
0 \in -\nabla_p G(\bar{x}, \bar{\mathbf{y}}, \bar{p}) + \mathcal{N}_{\mathcal{P}_k}(\bar{p}),
\end{cases}
\tag{26}
$$

then $(\bar{x}, \bar{\mathbf{y}}, \bar{p})$ satisfies (25). The main considerations for (26) are that $\mathcal{Z}_\epsilon(\bar{\mathbf{y}})$ is not convex for given $\bar{\mathbf{y}}$ if q is a nonlinear function, and $\mathcal{Z}_\epsilon(\cdot)$ is not Lipschitz continuous in the sense of Hausdorff distance, which can lead to failing the following convergence analysis (see e.g. [42]). In view of these, we use in the sequel $\mathcal{N}_X(\bar{x})$ rather than $\widehat{\mathcal{N}}_{\mathcal{Z}_\epsilon(\bar{\mathbf{y}})}(\bar{x})$.

MPEC is generally difficult to deal with because its constraints fail to satisfy the standard Mangasarian-Fromovitz constraint qualification (originated from [23]) at any feasible point [43, Proposition 1.1]. We recall the definition of MPEC linear independence constraint qualification (MPEC-LICQ).

Definition 6 ([33]) We say that MPEC-LICQ holds at $\mathbf{y}$ for problem (23) if

$$\{\mathbf{e}_i : i \in \mathcal{I}_{0+}(\mathbf{y}) \cup \mathcal{I}_{00}(\mathbf{y})\} \cup \{\mathbf{M}_i^\top + \mathbf{e}_i : i \in \mathcal{I}_{+0}(\mathbf{y}) \cup \mathcal{I}_{00}(\mathbf{y})\}$$

is linearly independent, where $\mathbf{e}_i$ is the ith column of the $mk \times mk$ identify matrix and $\mathbf{M}_i$ is the ith row of the matrix $\mathbf{M}$.

In what follows, we give the stationarity of problem (P). To this end, we assume that for arbitrary probability distribution/measure $P \in \mathcal{P}$ there exists a corresponding density function $p : \Xi \to \mathbb{R}_+$ such that $P(\mathrm{d}\xi) = p(\xi)\mathrm{d}\xi$[1]. Denote by $\mathfrak{P}$ the collection of all density functions of $P \in \mathcal{P}$.

To define the block coordinatewise stationarity of problem (P), we first give the definition of stationary points for stochastic MPEC (see Definition 3 for the discrete version).

For fixed $(\bar{x}, \bar{P})$, we consider the following stochastic MPEC problem

$$\min_{y} \theta(\bar{x}) + h(\mathbb{E}_{\bar{P}}[f(\bar{x}, y(\xi), \xi)]) \tag{27}$$
$$\text{s.t. } 0 \leq y(\xi) \perp M(\xi)y(\xi) + q(\bar{x}, \xi) \geq 0 \text{ for a.e. } \xi \in \Xi.$$

Define the following index sets:

$$\begin{aligned}
\mathcal{I}_{+0}(y; \xi) &= \{i : y_i(\xi) > 0, (M(\xi)y(\xi) + q(\bar{x}, \xi))_i = 0\}, \\
\mathcal{I}_{0+}(y; \xi) &= \{i : y_i(\xi) = 0, (M(\xi)y(\xi) + q(\bar{x}, \xi))_i > 0\}, \\
\mathcal{I}_{00}(y; \xi) &= \{i : y_i(\xi) = 0, (M(\xi)y(\xi) + q(\bar{x}, \xi))_i = 0\}.
\end{aligned} \tag{28}$$

Denote by $\mathcal{L}(\mathbb{R}^m)$ the collection of all measurable mappings from Ξ to $\mathbb{R}^m$.

Definition 7 (stationary points of problem (27)) (i) We say that $\bar{y}$ is a W-stationary point of problem (27) if there exist multipliers $(\lambda, \mu) \in \mathcal{L}(\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^m)$ such that for a.e. $\xi \in \Xi$,

$$\begin{aligned}
&\nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)])\nabla_y f(\bar{x}, \bar{y}(\xi), \xi) - \lambda(\xi) - M(\xi)^\top \mu(\xi) = 0, \\
&\lambda_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{+0}(\bar{y}; \xi), \ \mu_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{0+}(\bar{y}; \xi).
\end{aligned} \tag{29}$$

(ii) We say that $\bar{y}$ is a C-stationary point of problem (27) if there exist $(\lambda, \mu) \in \mathcal{L}(\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^m)$ satisfying (29) and for a.e. $\xi \in \Xi$,

$$\lambda_i(\xi)\mu_i(\xi) \geq 0 \text{ for } i \in \mathcal{I}_{00}(\bar{y}; \xi).$$

---

[1] We can generally assume that $P(\mathrm{d}\xi) = p(\xi)\mathbb{Q}(\mathrm{d}\xi)$ for some nominal probability distribution $\mathbb{Q}$. We know from Radon-Nikodym theorem (see e.g. [34, Theorem 7.32]) that there exists such a density function $p(\xi)$ if and only if P is absolutely continuous w.r.t. $\mathbb{Q}$. Here we neglect $\mathbb{Q}$ to simplify the notation.

(iii) We say that $\bar{y}$ is an M-stationary point of problem (27) if there exist $(\lambda, \mu) \in \mathcal{L}(\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^m)$ satisfying (29) and for a.e. $\xi \in \Xi$,

$$\lambda_i(\xi) > 0, \ \mu_i(\xi) > 0 \text{ or } \lambda_i(\xi)\mu_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{00}(\bar{y}; \xi).$$

(iv) We say that $\bar{y}$ is an S-stationary point of problem (27) if there exist $(\lambda, \mu) \in \mathcal{L}(\mathbb{R}^m) \times \mathcal{L}(\mathbb{R}^m)$ satisfying (29) and for a.e. $\xi \in \Xi$,

$$\lambda_i(\xi) \geq 0, \ \mu_i(\xi) \geq 0 \text{ for } i \in \mathcal{I}_{00}(\bar{y}; \xi).$$

Note that when Ξ has a finite number of elements, Definition 7 reduces to Definition 3.

Definition 8 (block coordinatewise •-stationary points of (P)**)** $(\bar{x}, \bar{y}, \bar{P})$ is called a block coordinatewise •-stationary point of problem (P) if it satisfies

$$\begin{cases} 0 \in \nabla_x \theta(\bar{x}) + \nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]) \mathbb{E}_{\bar{P}}[\nabla_x f(\bar{x}, \bar{y}(\xi), \xi)] + \mathcal{N}_X(\bar{x}), \\ \bar{y} \text{ is a •-stationary point of problem (27)}, \\ 0 \in -\nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]) f(\bar{x}, \bar{y}(\cdot), \cdot) + \mathcal{N}_{\mathfrak{P}}(\bar{p}), \end{cases}$$

where "•" can be W, C, M and S, $\bar{p}$ is the density function of $\bar{P}$ and $\mathcal{N}_{\mathfrak{P}}(\bar{p}) := \{v \in \mathcal{L}(\mathbb{R}) : \int_\Xi v(\xi)(p(\xi) - \bar{p}) \mathrm{d}\xi \leq 0, \forall p \in \mathfrak{P}\}$.

4.2 Convergence analysis

In this subsection, we study the convergence of block coordinatewise stationary points of $(\mathrm{P}_{\epsilon,k})$ defined by (26). We first consider the convergence of the block coordinatewise stationary points as $\epsilon \downarrow 0$ for a fixed k in the following theorem.

Theorem 7 *Let $(x_\epsilon^k, \mathbf{y}_\epsilon^k, p_\epsilon^k)$ be a block coordinatewise stationary point of problem $(\mathrm{P}_{\epsilon,k})$ defined by (26) and $(x^{k*}, \mathbf{y}^{k*}, p^{k*})$ be an accumulation point of $(x_\epsilon^k, \mathbf{y}_\epsilon^k, p_\epsilon^k)$ as $\epsilon \downarrow 0$. Suppose further that MPEC-LICQ holds at $\mathbf{y}^{k*}$ for problem (23) with $(\bar{x}, \bar{p}) = (x^{k*}, p^{k*})$ and $\epsilon = 0$. Then $(x^{k*}, \mathbf{y}^{k*}, p^{k*})$ is a block coordinatewise C-stationary point of problem (P_k).*

The proof of Theorem 7 is given in Appendix.

Let $\Xi^k := \{\xi^1, \cdots, \xi^k\}$ and its Voronoi tessellation be $\Xi_1, \cdots, \Xi_k$. For any feasible point of (P_k), denoted by $(x^k, \mathbf{y}^k, p^k)$, we make the following notations. Define the following density function:

$$\mathbf{p}^k(\xi) = \sum_{i=1}^k \frac{p_i^k}{\int_{\Xi_i} 1 \mathrm{d}\xi} \mathbf{1}_{\Xi_i}(\xi) \text{ for } \xi \in \Xi,$$

where $p^k = (p_1^k, \cdots, p_k^k)^\top$, and denote

$$\mathfrak{P}_k = \left\{ \mathbf{p}^k(\cdot) = \sum_{i=1}^k \frac{p_i^k}{\int_{\Xi_i} 1 \mathrm{d}\xi} \mathbf{1}_{\Xi_i}(\cdot) : p^k \in \mathcal{P}_k \right\}.$$

Denote by

$$y^k(\cdot) = \sum_{i=1}^{k} y^k(\xi^i) \mathbf{1}_{\Xi_i}(\cdot),$$

where $y^k(\xi^i)$ is the ith block of $\mathbf{y}^k = (y^k(\xi^1)^\top, \cdots, y^k(\xi^k)^\top)^\top$ for $i = 1, \cdots, k$. Denote by

$$\overline{M}_k(\cdot) = \sum_{i=1}^{k} M(\xi^i) \mathbf{1}_{\Xi_i}(\cdot) \text{ and } \overline{q}_k(x, \cdot) = \sum_{i=1}^{k} q(x, \xi^i) \mathbf{1}_{\Xi_i}(\cdot).$$

If, further, $(x^k, \mathbf{y}^k, p^k)$ is a block coordinatewise C-stationary point of problem (P_k), according to the definition of block coordinatewise C-stationary point, we have

$$\begin{cases} 0 \in \nabla_x G(x^k, \mathbf{y}^k, p^k) + \mathcal{N}_X(x^k), \\ \begin{cases} \nabla_{\mathbf{y}} G(x^k, \mathbf{y}^k, p^k) - \lambda^k - \mathbf{M}^\top \mu^k = 0, \\ \lambda_i^k = 0 \text{ for } i \in \mathcal{I}_{+0}(\mathbf{y}^k), \ \mu_i^k = 0 \text{ for } i \in \mathcal{I}_{0+}(\mathbf{y}^k), \\ \lambda_i^k \mu_i^k \geq 0 \text{ for } i \in \mathcal{I}_{00}(\mathbf{y}^k), \end{cases} \\ 0 \in -\nabla_p G(x^k, \mathbf{y}^k, p^k) + \mathcal{N}_{\mathcal{P}_k}(p^k), \end{cases} \tag{30}$$

where

$$\begin{aligned} \mathcal{I}_{+0}(\mathbf{y}^k) &= \{i : \mathbf{y}_i^k > 0, (\mathbf{M}\mathbf{y}^k + \mathbf{q}(x^k))_i = 0\}, \\ \mathcal{I}_{0+}(\mathbf{y}^k) &= \{i : \mathbf{y}_i^k = 0, (\mathbf{M}\mathbf{y}^k + \mathbf{q}(x^k))_i > 0\}, \\ \mathcal{I}_{00}(\mathbf{y}^k) &= \{i : \mathbf{y}_i^k = 0, (\mathbf{M}\mathbf{y}^k + \mathbf{q}(x^k))_i = 0\}. \end{aligned}$$

Denote by

$$\lambda^k(\cdot) = \sum_{i=1}^{k} \lambda^k(\xi^i) \mathbf{1}_{\Xi_i}(\cdot) \text{ and } \mu^k(\cdot) = \sum_{i=1}^{k} \mu^k(\xi^i) \mathbf{1}_{\Xi_i}(\cdot),$$

where $\lambda^k(\xi^i)$ is the ith block of $\lambda^k = (\lambda^k(\xi^1)^\top, \cdots, \lambda^k(\xi^k)^\top)^\top$ and $\mu^k(\xi^i)$ is the ith block of $\mu^k = (\mu^k(\xi^1)^\top, \cdots, \mu^k(\xi^k)^\top)^\top$ for $i = 1, \cdots, k$.

Then, we have the reformulation of (30) as follows: for every $\xi \in \Xi$,

$$\begin{cases} 0 \in \nabla_x \theta(x^k) + \nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) \mathbb{E}_{P^k}[\nabla_x f(x^k, y^k(\xi), \xi)] + \mathcal{N}_X(x^k), \\ \begin{cases} \nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) \nabla_y f(x^k, y^k(\xi), \xi) - \lambda^k(\xi) - \overline{M}_k(\xi)^\top \mu^k(\xi) = 0, \\ \lambda_i^k(\xi) = 0 \text{ for } i \in \mathcal{I}_{+0}(y^k; \xi), \ \mu_i^k(\xi) = 0 \text{ for } i \in \mathcal{I}_{0+}(y^k; \xi), \\ \lambda_i^k(\xi) \mu_i^k(\xi) \geq 0 \text{ for } i \in \mathcal{I}_{00}(y^k; \xi), \end{cases} \\ 0 \in -\nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) f(x^k, y^k(\cdot), \cdot) + \mathcal{N}_{\mathfrak{P}_k}(\mathbf{p}^k), \end{cases} \tag{31}$$

where P^k is the probability distribution of density function $\mathbf{p}^k$ and

$$\begin{aligned} \mathcal{I}_{+0}(y^k; \xi) &= \{i : y_i^k(\xi) > 0, (\overline{M}_k(\xi) y^k(\xi) + \overline{q}_k(x^k, \xi))_i = 0\}, \\ \mathcal{I}_{0+}(y^k; \xi) &= \{i : y_i^k(\xi) = 0, (\overline{M}_k(\xi) y^k(\xi) + \overline{q}_k(x^k, \xi))_i > 0\}, \\ \mathcal{I}_{00}(y^k; \xi) &= \{i : y_i^k(\xi) = 0, (\overline{M}_k(\xi) y^k(\xi) + \overline{q}_k(x^k, \xi))_i = 0\}. \end{aligned}$$

For $y_1, y_2 \in \mathcal{L}(\mathbb{R}^m)$, define the inner product and its induced norm by

$$\langle y_1, y_2 \rangle = \int_\Xi y_1(\xi)^\top y_2(\xi) \mathrm{d}\xi$$

and

$$\|y_1 - y_2\|_{\mathcal{L}_2} = \left(\int_\Xi \|y_1(\xi) - y_2(\xi)\|^2 \, \mathrm{d}\xi \right)^{\frac{1}{2}}.$$

Based on $\mathcal{L}_2$-norm, we can define the convergence relationship, denoted by $\xrightarrow{\mathcal{L}_2}$, and the deviation distance, denoted by $\mathrm{d}_{\mathcal{L}_2}(\cdot, \cdot)$.

The following theorem claims that: under certain conditions, a sequence of C-stationary points of problem (P_k) converges to block coordinatewise C-stationary points of problem (P) as $k \to \infty$.

Theorem 8 *Let $\{(x^k, \mathbf{y}^k, p^k)\}$ be a sequence of block coordinatewise C-stationary points of problem (P_k). Suppose that: (i) $x^k \to \bar{x}$, $y^k \xrightarrow{\mathcal{L}_2} \bar{y}$, $\mathbf{p}^k \xrightarrow{\mathcal{L}_2} \bar{\mathbf{p}}$ and $\mu^k \xrightarrow{\mathcal{L}_2} \bar{\mu}$ as $k \to \infty$; (ii) there exists $\kappa : \Xi \to \mathbb{R}_+$ satisfying $\int_\Xi \kappa(\xi)^2 \mathrm{d}\xi < \infty$, such that $\|f(x^k, y^k(\xi), \xi)\| \le \kappa(\xi)$, $\|f(\bar{x}, \bar{y}(\xi), \xi)\| \le \kappa(\xi)$, $\|\nabla_x f(x^k, y^k(\xi), \xi)\| \le \kappa(\xi)$, $\|\nabla_x f(\bar{x}, \bar{y}(\xi), \xi)\| \le \kappa(\xi)$, $|\mathbf{p}^k(\xi)| \le \kappa(\xi)$ and $|\bar{\mathbf{p}}(\xi)| \le \kappa(\xi)$ for a.e. $\xi \in \Xi$; (iii) $\beta_k \to 0$ as $k \to \infty$ where β_k is defined in (7); (iv) $\mathrm{d}_{\mathcal{L}_2}(\mathfrak{P}, \mathfrak{P}_k) \to 0$ as $k \to \infty$. Then $(\bar{x}, \bar{y}, \bar{\mathbf{p}})$ is a block coordinatewise C-stationary point of problem (P).*

Proof Note that

$$\left\| \mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)] - \mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)] \right\|$$

$$= \left\| \int_\Xi f(x^k, y^k(\xi), \xi) \mathbf{p}^k(\xi) \mathrm{d}\xi - \int_\Xi f(\bar{x}, \bar{y}(\xi), \xi) \bar{\mathbf{p}}(\xi) \mathrm{d}\xi \right\|$$

$$\le \left\| \int_\Xi (f(x^k, y^k(\xi), \xi) - f(\bar{x}, \bar{y}(\xi), \xi)) \mathbf{p}^k(\xi) \mathrm{d}\xi \right\|$$

$$+ \left\| \int_\Xi f(\bar{x}, \bar{y}(\xi), \xi)(\mathbf{p}^k(\xi) - \bar{\mathbf{p}}(\xi)) \mathrm{d}\xi \right\|.$$

Since $y^k \xrightarrow{\mathcal{L}_2} \bar{y}$ and $\mathbf{p}^k \xrightarrow{\mathcal{L}_2} \bar{\mathbf{p}}$, we have $y^k(\xi) \to \bar{y}(\xi)$ and $\mathbf{p}^k(\xi) \to \bar{\mathbf{p}}(\xi)$ for a.e. $\xi \in \Xi$ as $k \to \infty$. By the continuity of f, we have

$$\left\| f(x^k, y^k(\xi), \xi) - f(\bar{x}, \bar{y}(\xi), \xi) \right\| \left| \mathbf{p}^k(\xi) \right| \to 0$$

for a.e. $\xi \in \Xi$ as $k \to \infty$. Moreover,

$$\left\| f(x^k, y^k(\xi), \xi) - f(\bar{x}, \bar{y}(\xi), \xi) \right\| \left| \mathbf{p}^k(\xi) \right| \le 2\kappa^2(\xi).$$

By Lebesgue's dominated convergence theorem, we have

$$\left\| \int_\Xi (f(x^k, y^k(\xi), \xi) - f(\bar{x}, \bar{y}(\xi), \xi)) \mathbf{p}^k(\xi) \mathrm{d}\xi \right\|$$

$$\le \int_\Xi \left\| f(x^k, y^k(\xi), \xi) - f(\bar{x}, \bar{y}(\xi), \xi) \right\| \left| \mathbf{p}^k(\xi) \right| \mathrm{d}\xi \to 0$$

as $k \to \infty$, and similarly

$$\left\| \int_\Xi f(\bar{x}, \bar{y}(\xi), \xi)(\mathbf{p}^k(\xi) - \bar{\mathbf{p}}(\xi)) \mathrm{d}\xi \right\| \leq \int_\Xi \|f(\bar{x}, \bar{y}(\xi), \xi)\| \left|\mathbf{p}^k(\xi) - \bar{\mathbf{p}}(\xi)\right| \mathrm{d}\xi$$
$$\to 0$$

as $k \to \infty$. Therefore, $\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)] \to \mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]$ as $k \to \infty$. Due to the continuous differentiability of h, we obtain

$$\nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) \to \nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]) \text{ as } k \to \infty. \qquad (32)$$

Analogously, we have

$$\left\| \mathbb{E}_{P^k}[\nabla_x f(x^k, y^k(\xi), \xi)] - \mathbb{E}_{\bar{P}}[\nabla_x f(\bar{x}, \bar{y}(\xi), \xi)] \right\|$$
$$= \left\| \int_\Xi \nabla_x f(x^k, y^k(\xi), \xi)\mathbf{p}^k(\xi)\mathrm{d}\xi - \int_\Xi \nabla_x f(\bar{x}, \bar{y}(\xi), \xi)\bar{\mathbf{p}}(\xi)\mathrm{d}\xi \right\|$$
$$\leq \left\| \int_\Xi (\nabla_x f(x^k, y^k(\xi), \xi) - \nabla_x f(\bar{x}, \bar{y}(\xi), \xi))\mathbf{p}^k(\xi)\mathrm{d}\xi \right\|$$
$$+ \left\| \int_\Xi \nabla_x f(\bar{x}, \bar{y}(\xi), \xi)(\mathbf{p}^k(\xi) - \bar{\mathbf{p}}(\xi))\mathrm{d}\xi \right\|$$
$$\to 0$$

as $k \to \infty$. We obtain

$$\mathbb{E}_{P^k}[\nabla_x f(x^k, y^k(\xi), \xi)] \to \mathbb{E}_{\bar{P}}[\nabla_x f(\bar{x}, \bar{y}(\xi), \xi)] \text{ as } k \to \infty. \qquad (33)$$

Thus, by letting $k \to \infty$, we obtain from (32), (33) and the first equation of (31) that

$$0 \in \nabla_x \theta(\bar{x}) + \nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)])\mathbb{E}_{\bar{P}}[\nabla_x f(\bar{x}, \bar{y}(\xi), \xi)] + \mathcal{N}_X(\bar{x}). \qquad (34)$$

Since $M(\cdot), q(\cdot, \cdot)$ are continuous and $\beta_k \to 0$ as $k \to \infty$, $\overline{M}_k(\xi) \to M(\xi)$ and $\bar{q}_k(x^k, \xi) \to q(\bar{x}, \xi)$ as $k \to \infty$. By directly letting $k \to \infty$, we obtain from the second part of (31) that

$$\nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)])\nabla_y f(\bar{x}, \bar{y}(\xi), \xi) - \bar{\lambda}(\xi) - M(\xi)^\top \bar{\mu}(\xi) = 0$$

for a.e. $\xi \in \Xi$. Here $\bar{\lambda}(\xi)$ is the limit of $\{\lambda^k(\xi)\}$ in the sense of $\mathcal{L}_2$-norm as $k \to \infty$. Its existence is due to the convergence of sequences $\{x^k\}$, $\{y^k\}$ and $\{\mu^k\}$ as $k \to \infty$. Let $\mathcal{I}_{+0}(\bar{y}; \xi)$, $\mathcal{I}_{0+}(\bar{y}; \xi)$ and $\mathcal{I}_{00}(\bar{y}; \xi)$ be denoted in (28). Obviously, we have that for a.e. $\xi \in \Xi$, $\mathcal{I}_{+0}(\bar{y}; \xi) \subseteq \mathcal{I}_{+0}(y^k; \xi)$ and $\mathcal{I}_{0+}(\bar{y}; \xi) \subseteq \mathcal{I}_{0+}(y^k; \xi)$ for sufficiently large k. Thus, we have

$$\bar{\lambda}_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{+0}(\bar{y}; \xi) \text{ and } \bar{\mu}_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{0+}(\bar{y}; \xi).$$

By a similar discussion as the proof of Theorem 7, we obtain $\bar{\lambda}_i(\xi)\bar{\mu}_i(\xi) \geq 0$ for $i \in \mathcal{I}_{00}(\bar{y}; \xi)$. To sum up, we obtain

$$\begin{cases} \nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)])\nabla_y f(\bar{x}, \bar{y}(\xi), \xi) - \bar{\lambda}(\xi) - M(\xi)^\top \bar{\mu}(\xi) = 0, \\ \bar{\lambda}_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{+0}(\bar{y}; \xi), \ \bar{\mu}_i(\xi) = 0 \text{ for } i \in \mathcal{I}_{0+}(\bar{y}; \xi), \\ \bar{\lambda}_i(\xi)\bar{\mu}_i(\xi) \geq 0 \text{ for } i \in \mathcal{I}_{00}(\bar{y}(\xi); \xi). \end{cases} \qquad (35)$$

Since $\mathrm{d}_{\mathcal{L}_2}(\mathfrak{P}, \mathfrak{P}_k) \to 0$ as $k \to \infty$, for any $\rho \in \mathfrak{P}$, there exists $\rho^k \in \mathfrak{P}_k$ such that $\rho^k \xrightarrow{\mathcal{L}_2} \rho$ as $k \to \infty$. Thus, we have

$$\rho^k - \mathbf{p}^k \xrightarrow{\mathcal{L}_2} \rho - \bar{\mathbf{p}} \text{ and } f(x^k, y^k(\cdot), \cdot) \xrightarrow{\mathcal{L}_2} f(\bar{x}, \bar{y}(\cdot), \cdot) \tag{36}$$

as $k \to \infty$. Due to

$$0 \in -\nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) f(x^k, y^k(\cdot), \cdot) + \mathcal{N}_{\mathfrak{P}_k}(\mathbf{p}^k),$$

we have $\langle \nabla h(\mathbb{E}_{P^k}[f(x^k, y^k(\xi), \xi)]) f(x^k, y^k(\cdot), \cdot), \rho^k - \mathbf{p}^k \rangle \leq 0$. Based on (32) and (36), by letting $k \to \infty$, we obtain

$$\langle \nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]) f(\bar{x}, \bar{y}(\cdot), \cdot), \rho - \bar{\mathbf{p}} \rangle \leq 0$$

for any $\rho \in \mathfrak{P}$, which is equivalent to

$$0 \in -\nabla h(\mathbb{E}_{\bar{P}}[f(\bar{x}, \bar{y}(\xi), \xi)]) f(\bar{x}, \bar{y}(\cdot), \cdot) + \mathcal{N}_{\mathfrak{P}}(\bar{\mathbf{p}}). \tag{37}$$

Finally, combining (34) and (35) with (37), we complete the proof. □

Remark 3 Assumption (iv) in Theorem 8 is used to derive (37). In fact, to have (37), we only need that for any $\rho \in \mathfrak{P}$, there is a sequence $\{\rho^k\} \subseteq \mathfrak{P}_k$ such that $\rho^k \xrightarrow{\mathcal{L}_2} \rho$ as $k \to \infty$. If Γ is an open set and $\beta_k \to 0$ (defined in (7)) as $k \to \infty$, then we can always find such ρ^k due to the integrability of ρ and the definition of Riemann integral. In this case, (37) holds automatically.

The limit of $\mathcal{N}_{\mathcal{P}_k}(p^k)$ is unclear because the dimension of $\mathcal{P}_k$ is different with different sample size k. On the other hand, if $\mathrm{d}_{\mathcal{L}_2}(\mathfrak{P}, \mathfrak{P}^k) \to 0$ and $\mathbf{p}^k \xrightarrow{\mathcal{L}_2} \bar{\mathbf{p}}$ as $k \to \infty$, we know from the proof of Theorem 8 that $\limsup_{k \to \infty} \mathcal{N}_{\mathfrak{P}^k}(\mathbf{p}^k) \subseteq \mathcal{N}_{\mathfrak{P}}(\bar{\mathbf{p}})$.

At the end of this section, we provide a diagram in Figure 1 that summarizes the convergence relationships between problems (P), (P_ϵ), (P_k) and ($P_{\epsilon,k}$) as $\epsilon \downarrow 0$ and $k \to \infty$.
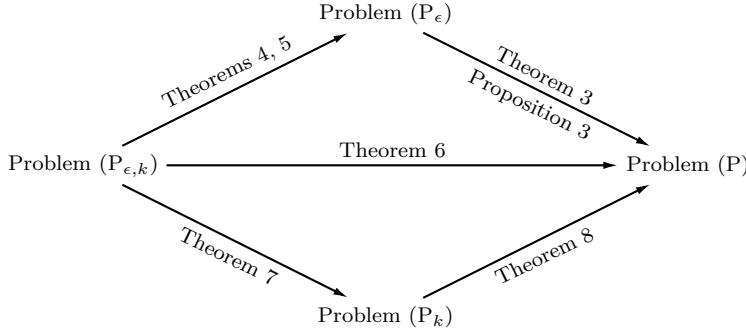
Fig. 1 Convergence relationships between problems (P), (P_ϵ), (P_k) and ($P_{\epsilon,k}$) as $\epsilon \downarrow 0$ and $k \to \infty$

5 Pure characteristics demand models

In this section, we apply our results in the previous sections to the pure characteristics demand model when the underlying probability distribution is uncertain. In subsection 5.1, we present an alternating algorithm to solve problem $(\mathrm{P}_{\epsilon,k})$ under a general setting. In subsection 5.2, we report numerical results for some concrete examples.

5.1 An alternating algorithm for problem $(\mathrm{P}_{\epsilon,k})$

Problem $(\mathrm{P}_{\epsilon,k})$ can be rewritten as (see (10))

$$\min_{x \in X} \ \theta(x) + \max_{p \in \mathcal{P}_k} h(F(x, \hat{\mathbf{y}}_\epsilon(x))p). \tag{38}$$

The following algorithm gives an alternating algorithm to solve (38).

Algorithm 1 *Choose an initial point $p^0 \in \mathcal{P}_k$. Let $j = 0$.*
  Step 1: *Generate x^j by solving*

$$\min_{x \in X} \ \theta(x) + h(F(x, \hat{\mathbf{y}}_\epsilon(x))p^j).$$

  Step 2: *Generate p^{j+1} by solving*

$$\max_{p \in \mathcal{P}_k} \ h(F(x^j, \hat{\mathbf{y}}_\epsilon(x^j))p).$$

  *Let $j = j + 1$ and go to **Step 1**.*

Alternating algorithms are widely-used to solve minimax problems (see e.g. [14, 40] and the references therein). Suppose the feasible set X is bounded. The minimization problem in Step 1 of Algorithm 1 has a minimizer for any $p^j \in \mathcal{P}_k$, since $\hat{\mathbf{y}}_\epsilon(x)$ is well-defined for any $x \in X$ and the objective function is Lipschitz continuous. Moreover, $\hat{\mathbf{y}}_\epsilon(x)$ can be written explicitly as a function of x in a closed form [4]. We can use existing continuous optimization methods or codes to solve the minimization problem. The maximization problem in Step 2 has a maximizer for any x^j, since h is a continuous function w.r.t. p and the feasible set $\mathcal{P}_k$ is compact and convex. If h is convex (see the specific settings in numerical parts), we can use some existing algorithms for difference-of-convex optimization problems to solve the maximization problem.

It is noteworthy that the Wasserstein metric can be used to construct a data-driven ambiguity set [10]. However, the data-driven Wasserstein ball ambiguity set in [10] cannot be directly applied to solve problem (P). In this paper, we adopt a general moment information ambiguity set, and then use its discrete approximation to approximately solve the original problem. Our discrete approximation problem (38) of the ambiguity set has a finite support set. The data-driven Wasserstein ball ambiguity set in [10] still has the same support set as the original Wasserstein ball, which just replaces the nominal probability distribution P by its data-driven counterpart P_N, i.e., $\mathcal{P} := \{Q \in \mathcal{P}(\Xi) : \mathbb{D}_W(Q, P) \leq r\}$ and $\mathcal{P}_N := \{Q \in \mathcal{P}(\Xi) : \mathbb{D}_W(Q, P_N) \leq r\}$ for some $r > 0$.

5.2 Numerical results

In particular, we consider the distributionally robust counterpart of problem
(4):

$$\min_{x \in X, s_t} \frac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle + \varrho \max_{P \in \mathcal{P}} \sum_{t=1}^{T} \| A_t \mathbb{E}_P[s_t(\xi)] - b_t \|^2$$
$$\text{s.t.} \quad 0 \leq z_t(\xi) \perp M z_t(\xi) + q_t(x, \xi) \geq 0, \quad t = 1, 2, \cdots, T, \tag{39}$$

where $\mathcal{P} := \{ P \in \mathcal{P}(\Xi) : (\mathbb{E}_P[\xi] - \mu_0)^\top (\mathbb{E}_P[\xi] - \mu_0) \leq \eta \}$, $\mu_0 \in \mathbb{R}^\nu$ and
$\eta > 0$, which assumes that the mean of ξ lies in a ball of size η centered at the
estimate μ_0 (see e.g. [9, (1a)]). By employing the Schur complement, we can
reformulate $\mathcal{P}$ as the form (5):

$$\mathcal{P} := \left\{ P \in \mathcal{P}(\Xi) : \mathbb{E}_P \left[\begin{pmatrix} I & \xi - \mu_0 \\ (\xi - \mu_0)^\top & \eta \end{pmatrix} \right] \in \Gamma \right\},$$

where the cone Γ denotes the set of positive semidefinite $(\nu + 1) \times (\nu + 1)$
matrices. Here, we use the first-moment information as an ambiguity set, which
can be captured intuitively in practice and its modeling motivation is natural
(see e.g. [12]). Moreover, we can use the first-moment information ambiguity
set to illustrate our theoretical results with numerical examples clearly.

The discretization problem of (39) is

$$\min_{x \in X, s_t} \frac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle + \varrho \max_{p \in \mathcal{P}_k} \sum_{t=1}^{T} \left\| A_t \sum_{i=1}^{k} s_t(\xi^i) p_i - b_t \right\|^2$$
$$\text{s.t.} \quad 0 \leq z_t(\xi^i) \perp M z_t(\xi^i) + q_t(x, \xi^i) \geq 0, \ t = 1, 2, \cdots, T, \ i = 1, 2 \cdots, k,$$
$$\tag{40}$$

where $\{\xi^1, \cdots, \xi^k\} \subseteq \Xi$ and

$$\mathcal{P}_k := \left\{ p \in \mathbb{R}_+^k : \sum_{i=1}^{k} p_i = 1, \ \left\| \sum_{i=1}^{k} p_i \xi^i - \mu_0 \right\|^2 \leq \eta \right\}.$$

Then the regularization and discretization problem is

$$\min_{x \in X} \frac{1}{2} \langle x, Hx \rangle + \langle c, x \rangle + \varrho \max_{p \in \mathcal{P}_k} \sum_{t=1}^{T} \left\| A_t \sum_{i=1}^{k} s_{t,\epsilon}(x, \xi^i) p_i - b_t \right\|^2,$$

where $z_{t,\epsilon}(x, \xi^i) \in \text{SOL}(M + \epsilon I, q_t(x, \xi^i))$, $z_{t,\epsilon}(x, \xi^i) := (s_{t,\epsilon}(x, \xi^i)^\top, \gamma_{t,\epsilon}(x, \xi^i))^\top$
for $i = 1, \cdots, k$, $t = 1, \cdots, T$.

We adopt the following settings, which follow from [4, Example 4.1]. The
utility function in market t is given by

$$u_t(x, \xi) = \mathbf{C}_t \chi_1(x_2, x_3, \xi_1) - \chi_2(x_4, \xi_2) \sigma_t + x_{1t},$$

where $\mathbf{C}_t = (C_{1t}, \cdots, C_{\tau t}) \in \mathbb{R}^{m \times \tau}$, $C_{jt} \in \mathbb{R}^m$, $x_{1t} \in \mathbb{R}^m$, $x_2, x_3 \in \mathbb{R}^\tau$,
$x_4 \in \mathbb{R}$. Let $\mathbf{x}_1 = (x_{11}^\top, \cdots, x_{1T}^\top)^\top \in \mathbb{R}^{mT}$, $\mathbf{x}_2 = (x_2^\top, x_3^\top, x_4)^\top$ and $x = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top \in \mathbb{R}^n$ with $n = mT + 2\tau + 1$, $\xi = (\xi_1, \xi_2)^\top : \Omega \to \Xi \subseteq \mathbb{R}^2$,
$\chi_1(x_2, x_3, \xi_1) = x_2 + x_3 \xi_1$, $\chi_2(x_4, \xi_2) = \exp(x_4 \xi_2)$, $\sigma_t \in \mathbb{R}^m$.

To numerically present these convergence results in Section 3, we make the following specific settings. Set $T = 1$, $m = 2$, $\tau = 1$, $n = 5$, $b_1 = (0.5, 0.5)^\top$, $C_1 = (2, 3)^\top$, $\sigma_1 = (1, 2)^\top$ and $X = [0, 2]^5$, $\Xi = [-1, 1]^2$, $\mu_0 = (0, 0)^\top$. Consider

$$\min_{x \in X} \max_{p \in \mathcal{P}_k} G(x, \hat{\mathbf{z}}_\epsilon(x), p) := \tfrac{1}{2} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle + \left\| \sum_{i=1}^k s_{1,\epsilon}(x, \xi^i) p_i - b_1 \right\|^2. \qquad (41)$$

Here $(s_{1,\epsilon}(x, \xi^i)^\top, \gamma_{1,\epsilon}(x, \xi^i))^\top = z_{1,\epsilon}(x, \xi^i) \in \mathrm{SOL}(M + \epsilon I, q_1(x, \xi^i))$, $i = 1, \cdots, k$ and $\hat{\mathbf{z}}_\epsilon(x) := (z_{1,\epsilon}(x, \xi^1)^\top, \cdots, z_{1,\epsilon}(x, \xi^k)^\top)^\top$.

It is easy to check that $x^* = (0, 0, 1, 0, 0)^\top$ is an optimal solution of problem (41) and the corresponding optimal value is 0. To see this, we have

$$\chi_1(x_2^*, x_3^*, \xi_1) = x_2^* + x_3^* \xi_1 = 1 \text{ and } \chi_2(x_4^*, \xi_2) = \exp(x_4^* \xi_2) = 1.$$

Therefore, $u_1(x^*, \xi) = C_1 \chi_1(x_2^*, x_3^*, \xi_1) - \chi_2(x_4^*, \xi_2) \sigma_1 + \mathbf{x}_1^* = (1, 1)^\top$. According to (2), we know that the solution set of $s_1(\xi)$ is $\{(\varsigma, 1 - \varsigma)^\top : 0 \leq \varsigma \leq 1\}$, where the least norm solution is $(0.5, 0.5)^\top$ for every $\xi \in \Xi$. This implies that $\mathbb{E}_P[s_{1,\epsilon}(x^*, \xi)] = \frac{1}{2+\epsilon}(1 + \epsilon, 1 + \epsilon)^\top \to (0.5, 0.5)^\top$ as $\epsilon \downarrow 0$. Thus, the conditions in Proposition 3 hold.

Under the above settings, $\Psi(\xi) = \begin{pmatrix} I & \xi \\ \xi^\top & \eta \end{pmatrix}$ and Γ is the set of positive semidefinite 3×3 matrices. Assumption 1 holds with P_0 being the uniform distribution over $[-1, 1]^2$ and α being any positive scalar less than or equal to η. Moreover, Assumption 2 holds with $\kappa(\xi) = 1$. Therefore, the convergence results in Section 3 hold.

We use Algorithm 1 to report some numerical results of problem (41). That is, select an arbitrary initial point $p^0 \in \mathcal{P}_k$. For $j = 0, 1, 2, \cdots$, do

$$x^j \in \arg\min_{x \in X} \ G(x, \hat{\mathbf{z}}_\epsilon(x), p^j) \qquad (42)$$

and

$$p^{j+1} \in \arg\max_{p \in \mathcal{P}_k} \ G(x^j, \hat{\mathbf{z}}_\epsilon(x^j), p). \qquad (43)$$

Due to the special structure of matrix M, we adopt the closed-form solution in [4] to compute $\hat{\mathbf{z}}_\epsilon(x^j)$. The function $G(x^j, \hat{\mathbf{z}}_\epsilon(x^j), \cdot)$ is a quadratic convex function. We can use algorithm in [44] to find a maximizer of p^j in (43) on the bounded convex set $\mathcal{P}_k$. Since $\mathcal{P}_k$ and X are bounded, the sequence $\{(x^j, p^j)\}$ generated by Algorithm 1 has at least one accumulation point. We employ (26) as the stopping criterion. Actually, we only need to verify $0 \in \nabla_x G(x^j, \hat{\mathbf{z}}_\epsilon(x^j), p^j) + \mathcal{N}_X(x^j)$. Due to the box structure of X, the projection onto X can be computed easily. Thus, we stop the iteration when

$$\left\| x^j - \mathrm{Proj}_X(x^j - \nabla_x G(x^j, \hat{\mathbf{z}}_\epsilon(x^j), p^j)) \right\| \leq 10^{-4}. \qquad (44)$$

We chose an initial point $p^0 \in \mathcal{P}_k$ with $p_i^0 = \frac{1}{k}, i = 1, \ldots, k$. First, for fixed $\eta = 0.1, 0.2, 0.5, 1$ and sample size $k = 25, 2500$, we compute optimal values of problem (41) w.r.t. $\epsilon = 0.5, 0.2, 0.1, 0.05, 0.01$. We present these results in

Figure 2. It shows the tendency that the optimal value of problem (41) tends to zero as ϵ goes to zero. Meanwhile, for fixed each $\epsilon = 0.5, 0.2, 0.1, 0.05, 0.01$, we can observe from Figure 2 that the optimal value of problem (41) increases as η increases, which shows that the distributionally robust model (41) works as expected.
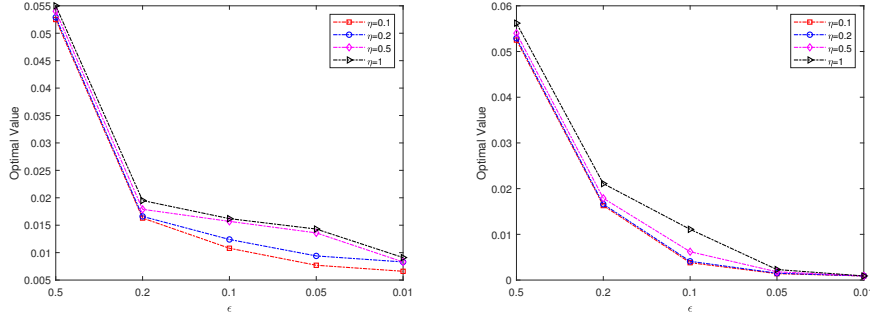
(a) Convergence of optimal values for $k = 25$ as ϵ decreases

(b) Convergence of optimal values for $k = 2500$ as ϵ decreases

Fig. 2 Convergence of optimal values as ϵ decreases

Furthermore, for fixed $\eta = 0.1, 0.2, 0.5, 1$ and $\epsilon = 0.5, 0.2$, we compute optimal values of problem (41) with different sample sizes, see Figure 3. It shows that, for each fixed ϵ, the optimal values converge as sample size goes to infinity. Moreover, we present in Table 1 the distances between x^j satisfying (44) and the true solution $(0, 0, 1, 0, 0)^\top$ with different ϵ and k for fixed $\eta = 0.5$. It shows the convergence of optimal solutions as sample size goes to infinity for fixed ϵ.
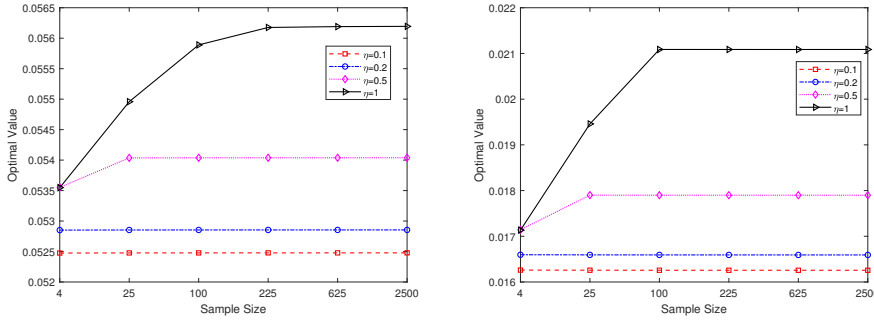
(a) Convergence of optimal values as sample size increases for $\epsilon = 0.5$

(b) Convergence of optimal values as sample size increases for $\epsilon = 0.2$

Fig. 3 Optimal values w.r.t. different sample sizes

Table 1 $\|x^j - x^*\|$ with different ϵ and k for fixed $\eta = 0.5$ where x^j satisfies (44) and $x^* = (0,0,1,0,0)^\top$

ϵ \ k	4	25	100	225	625	2500
0.5	0.5688	0.5210	0.6791	0.6887	0.7091	0.7202
0.2	0.7851	0.4422	0.3195	0.2215	0.2119	0.2102
0.1	0.6649	0.1079	0.0851	0.0853	0.0867	0.0856

6 Concluding remarks

This paper considers a class of distributionally robust mathematical programs with stochastic complementarity constraints (DRMP-SCC) in the form of problem (P), which arise from pure characteristics demand models under uncertainties of probability distributions of the involved random variables. Since problem (P) is a nonconvex-nonconcave minimax problem, minimax is not equal to maximin and thus a saddle point does not exist in general. We define global and local optimality and stationary points of problem (P), and its discretization and/or regularization approximation problems (P_ϵ), (P_k) and $(P_{\epsilon,k})$. We provide sufficient conditions for the convergence of optimal solutions and stationary points of problem $(P_{\epsilon,k})$ as ϵ goes to zero and k goes to infinity. We show that all those conditions hold for pure characteristics demand models under uncertainties. Moreover, we use numerical results to show the effectiveness of our theoretical results.

Appendix

Proof (The proof of Proposition 1)

Denote $\bar{p}_i = P_0(\Xi_i)$ for $i = 1, \cdots, k$. We verify that $\bar{p} = (\bar{p}_1, \cdots, \bar{p}_k)^\top \in \mathcal{P}_k$ for all sufficiently large k in the following. Since Ψ is continuous, we know from mean value theorem of integrals that

$$\mathbb{E}_{P_0}[\Psi(\xi)] = \sum_{i=1}^{k} \int_{\Xi_i} \Psi(\xi)\, P_0(d\xi) = \sum_{i=1}^{k} \Psi(\tilde{\xi}^i) P_0(\Xi_i)$$

for some $\tilde{\xi}^i \in \Xi_i$, $i = 1, \cdots, k$. Then

$$\left\| \mathbb{E}_{P_0}[\Psi(\xi)] - \sum_{i=1}^{k} \bar{p}_i \Psi(\xi^i) \right\| \leq \sum_{i=1}^{k} \bar{p}_i \left\| \Psi(\xi^i) - \Psi(\tilde{\xi}^i) \right\|. \tag{45}$$

We first consider the case that Ξ is bounded. For $\alpha > 0$, there exists $\delta > 0$ such that if $\max_{1 \leq i \leq k} \mathrm{diam}(\Xi_i) < \delta$, then

$$\max_{1 \leq i \leq k} \left\| \Psi(\xi^i) - \Psi(\tilde{\xi}^i) \right\| \leq \alpha.$$

Since Ξ is bounded, we can find a sequence $\{\xi^k\}_{k=1}^\infty$ such that the corresponding Voronoi tessellation $\Xi_1, \cdots, \Xi_k, \cdots$ satisfying

$$\lim_{k\to\infty} \max_{1\le i\le k} \operatorname{diam}(\Xi_i) = 0.$$

Hence there is $\bar{k} > 0$ such that $\max_{1\le i\le k} \operatorname{diam}(\Xi_i) < \delta$ for any $k \ge \bar{k}$.

Then, it knows from (45) that

$$\left\| \mathbb{E}_{P_0}[\Psi(\xi)] - \sum_{i=1}^k \bar{p}_i \Psi(\xi^i) \right\| \le \alpha.$$

This, together with Assumption 1, indicates that $\sum_{i=1}^k \bar{p}_i \Psi(\xi^i) \in \Gamma$, which implies the nonemptiness of $\mathcal{P}_k$.

Now we consider the case that Ξ is unbounded. Let $\Xi_b := \{\xi \in \Xi : \|\xi\| \le b\}$ for $b > 0$. Denote a probability distribution $\bar{P}_0$ supported on Ξ_b by

$$\bar{P}_0(\Xi_a) = \frac{P_0(\Xi_a \cap \Xi_b)}{P_0(\Xi_b)}$$

for any measurable $\Xi_a \subseteq \Xi$, where P_0 is defined in Assumption 1. Note that

$$\lim_{b\to\infty} \frac{1}{P_0(\Xi_b)} = 1 \text{ and } \lim_{b\to\infty} \int_{\Xi_b} \Psi(\xi) P_0(d\xi) = \int_\Xi \Psi(\xi) P_0(d\xi) = \mathbb{E}_{P_0}[\Psi(\xi)].$$

We have

$$\lim_{b\to\infty} \int_{\Xi_b} \Psi(\xi) \bar{P}_0(d\xi) = \lim_{b\to\infty} \frac{1}{P_0(\Xi_b)} \int_{\Xi_b} \Psi(\xi) P_0(d\xi) = \mathbb{E}_{P_0}[\Psi(\xi)].$$

Therefore, there exists $b_0 > 0$ such that, for any $b \ge b_0$,

$$\left\| \mathbb{E}_{\bar{P}_0}[\Psi(\xi)] - \mathbb{E}_{P_0}[\Psi(\xi)] \right\| \le \frac{\alpha}{2}.$$

From Assumption 1, we obtain

$$\mathbb{E}_{\bar{P}_0}[\Psi(\xi)] + \frac{\alpha}{2}\mathbb{B} \subseteq \Gamma. \tag{46}$$

Due to the boundedness of Ξ_b and (46), by the same proof for the case that Ξ is bounded, there exists a $\bar{k} > 0$ such that $\mathcal{P}_k$ is nonempty for $k \ge \bar{k}$. $\qquad\square$

Proof (The proof of Theorem 7) Since $(x^{k*}, \mathbf{y}^{k*}, p^{k*})$ is an accumulation point of $(x_\epsilon^k, \mathbf{y}_\epsilon^k, p_\epsilon^k)$ as $\epsilon \downarrow 0$, there exists a sequence $\{\epsilon_j\}_{j=1}^\infty$ with $\epsilon_j \downarrow 0$ as $j \to \infty$, such that $(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) \to (x^{k*}, \mathbf{y}^{k*}, p^{k*})$ as $j \to \infty$. Based on (26), we have

$$\begin{cases} 0 \in \nabla_x G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_X(x_{\epsilon_j}^k), \\ \mathbf{y}_{\epsilon_j}^k = \hat{\mathbf{y}}_{\epsilon_j}(x_{\epsilon_j}^k), \\ 0 \in -\nabla_p G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_{\mathcal{P}_k}(p_{\epsilon_j}^k). \end{cases}$$

We know from Definitions 3, 4 and Lemma 3 that

$$
\begin{cases}
0 \in \nabla_x G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_X(x_{\epsilon_j}^k), \\
\begin{cases}
\nabla_{\mathbf{y}} G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) - \lambda^{k,j} - (\mathbf{M}^\top + \epsilon_j I)\mu^{k,j} = 0, \\
\lambda_i^{k,j} = 0 \text{ for } i \in \mathcal{I}_{+0}^{k,j}(\mathbf{y}_{\epsilon_j}^k), \ \mu_i^{k,j} = 0 \text{ for } i \in \mathcal{I}_{0+}^{k,j}(\mathbf{y}_{\epsilon_j}^k), \\
\lambda_i^{k,j} \geq 0, \ \mu_i^{k,j} \geq 0 \text{ for } i \in \mathcal{I}_{00}^{k,j}(\mathbf{y}_{\epsilon_j}^k), \\
\end{cases} \\
0 \in -\nabla_p G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_{\mathcal{P}_k}(p_{\epsilon_j}^k),
\end{cases}
$$

where $\{\lambda^{k,j}\}_{j=1}^\infty$ and $\{\mu^{k,j}\}_{j=1}^\infty$ are two sequences of multipliers, and

$$
\begin{aligned}
\mathcal{I}_{+0}^{k,j}(\mathbf{y}_{\epsilon_j}^k) &= \{i : (\mathbf{y}_{\epsilon_j}^k)_i > 0, ((\mathbf{M} + \epsilon_j I)\mathbf{y}_{\epsilon_j}^k + \mathbf{q}(x_{\epsilon_j}^k))_i = 0\}, \\
\mathcal{I}_{0+}^{k,j}(\mathbf{y}_{\epsilon_j}^k) &= \{i : (\mathbf{y}_{\epsilon_j}^k)_i = 0, ((\mathbf{M} + \epsilon_j I)\mathbf{y}_{\epsilon_j}^k + \mathbf{q}(x_{\epsilon_j}^k))_i > 0\}, \\
\mathcal{I}_{00}^{k,j}(\mathbf{y}_{\epsilon_j}^k) &= \{i : (\mathbf{y}_{\epsilon_j}^k)_i = 0, ((\mathbf{M} + \epsilon_j I)\mathbf{y}_{\epsilon_j}^k + \mathbf{q}(x_{\epsilon_j}^k))_i = 0\}.
\end{aligned}
$$

Thus, for sufficiently large j, we have

$$
\begin{cases}
0 \in \nabla_x G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_X(x_{\epsilon_j}^k), \\
\begin{cases}
\nabla_{\mathbf{y}} G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) - \lambda^{k,j} - (\mathbf{M}^\top + \epsilon_j I)\mu^{k,j} = 0, \\
\lambda_i^{k,j} = 0 \text{ for } i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*}), \ \mu_i^{k,j} = 0 \text{ for } i \in \mathcal{I}_{0+}^k(\mathbf{y}^{k*}), \\
\end{cases} \\
0 \in -\nabla_p G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k) + \mathcal{N}_{\mathcal{P}_k}(p_{\epsilon_j}^k),
\end{cases} \tag{47}
$$

where

$$
\begin{aligned}
\mathcal{I}_{+0}(\mathbf{y}^{k*}) &= \{i : y_i^{k*} > 0, (\mathbf{M}\mathbf{y}^{k*} + \mathbf{q}(x^{k*}))_i = 0\}, \\
\mathcal{I}_{0+}(\mathbf{y}^{k*}) &= \{i : y_i^{k*} = 0, (\mathbf{M}\mathbf{y}^{k*} + \mathbf{q}(x^{k*}))_i > 0\}.
\end{aligned}
$$

Next, we verify the boundedness of sequences $\{\lambda^{k,j}\}_{j=1}^\infty$ and $\{\mu^{k,j}\}_{j=1}^\infty$. Notice that if $\{\mu^{k,j}\}_{j=1}^\infty$ is bounded, from the boundedness of $\{(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k)\}_{j=1}^\infty$ and continuous differentiability of G, $\{\lambda^{k,j}\}_{j=1}^\infty$ is bounded. Now we assume that $\{\mu^{k,j}\}_{j=1}^\infty$ is unbounded. Consider, by dividing $\|\mu^{k,j}\|$, that

$$
\frac{\nabla_{\mathbf{y}} G(x_{\epsilon_j}^k, \mathbf{y}_{\epsilon_j}^k, p_{\epsilon_j}^k)}{\|\mu^{k,j}\|} - \frac{\lambda^j}{\|\mu^{k,j}\|} - (\mathbf{M}^\top + \epsilon_j I)\frac{\mu^{k,j}}{\|\mu^{k,j}\|} = 0,
$$

which can deduce, according to $\|\mu^{k,j}\| \to \infty$ as $j \to \infty$, that

$$
\frac{\lambda^{k,j}}{\|\mu^{k,j}\|} + \mathbf{M}^\top \frac{\mu^{k,j}}{\|\mu^{k,j}\|} \to 0 \tag{48}
$$

as $j \to \infty$. Since $\lambda_i^{k,j} = 0$ for $i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*})$ and $\mu_i^{k,j} = 0$ for $i \in \mathcal{I}_{0+}^k(\mathbf{y}^{k*})$, we rewrite (48) as

$$
\sum_{i \in \mathcal{I}_{0+}^k(\mathbf{y}^{k*}) \cup \mathcal{I}_{00}^k(\mathbf{y}^{k*})} \frac{\lambda_i^{k,j}}{\|\mu^{k,j}\|} \mathbf{e}_i + \sum_{i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*}) \cup \mathcal{I}_{00}^k(\mathbf{y}^{k*})} \frac{\mu_i^{k,j}}{\|\mu^{k,j}\|} \mathbf{M}_i^\top \to 0,
$$

where $\mathcal{I}_{00}^k(\mathbf{y}^{k*}) = \{i : \mathbf{y}_i^{k*} = 0, (\mathbf{M}\mathbf{y}^{k*} + \mathbf{q}(x^{k*}))_i = 0\}$.

Then, by MPEC-LICQ at $\mathbf{y}^{k*}$ for problem (23) with $(\bar{x}, \bar{p}) = (x^{k*}, p^{k*})$ and $\epsilon = 0$, we obtain

$$\frac{\lambda_i^{k,j}}{\|\mu^{k,j}\|} \to 0, i \in \mathcal{I}_{0+}^k(\mathbf{y}^{k*}) \cup \mathcal{I}_{00}^k(\mathbf{y}^{k*}) \text{ and } \frac{\mu_i^{k,j}}{\|\mu^{k,j}\|} \to 0, i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*}) \cup \mathcal{I}_{00}^k(\mathbf{y}^{k*})$$

as $k \to \infty$, which contradicts $\frac{\mu_i^{k,j}}{\|\mu^{k,j}\|} \nrightarrow 0$ for some $i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*}) \cup \mathcal{I}_{00}^k(\mathbf{y}^{k*})$. Therefore, both $\{\lambda^{k,j}\}_{j=1}^\infty$ and $\{\mu^{k,j}\}_{j=1}^\infty$ are bounded. Without loss of generality, we assume that $\lambda^{k,j} \to \lambda^{k*}$ and $\mu^{k,j} \to \mu^{k*}$ as $j \to \infty$. Therefore, by letting $j \to \infty$, we have from (47) that

$$\begin{cases} 0 \in \nabla_x G(x^{k*}, \mathbf{y}^{k*}, p^{k*}) + \mathcal{N}_X(x^{k*}), \\ \begin{cases} \nabla_\mathbf{y} G(x^{k*}, \mathbf{y}^{k*}, p^{k*}) - \lambda^{k*} - \mathbf{M}^\top \mu^{k*} = 0, \\ \lambda_i^{k*} = 0 \text{ for } i \in \mathcal{I}_{+0}^k(\mathbf{y}^{k*}), \ \mu_i^{k*} = 0 \text{ for } i \in \mathcal{I}_{0+}^k(\mathbf{y}^{k*}), \end{cases} \\ 0 \in -\nabla_p G(x^{k*}, \mathbf{y}^{k*}, p^{k*}) + \mathcal{N}_{\mathcal{P}_k}(p^{k*}). \end{cases} \quad (49)$$

Moreover, for $\lambda_i^{k,j}\mu_i^{k,j} \geq 0$ for $i = 1, \ldots, mk$, we have $\lambda_i^{k*}\mu_i^{k*} \geq 0$ for $i \in \mathcal{I}_{00}^k(\mathbf{y}^{k*})$. This together with (49) means that $(x^{k*}, \mathbf{y}^{k*}, p^{k*})$ is a block coordinatewise C-stationary point of problem (P_k). $\square$

References

1. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. **12**, 1–12 (1965)
2. Berry, S., Pakes, A.: The pure characteristics demand model. Internat. Econom. Rev. **48**, 1193–1225 (2007)
3. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2013)
4. Chen, X., Sun, H., Wets, R.J.B.: Regularized mathematical programs with stochastic equilibrium constraints: Estimating structural demand models. SIAM J. Optim. **25**, 53–75 (2015)
5. Chen, X., Xiang, S.: Perturbation bounds of P-matrix linear complementarity problems. SIAM J. Optim. **18**, 1250–1265 (2008)
6. Chen, X., Ye, J.: A class of quadratic programs with linear complementarity constraints. Set-Valued Var. Anal. **17**, 113–133 (2009)
7. Cottle, R.W., Pang, J.S., Stone, R.E.: The Linear Complementarity Problem. SIAM, Philadelphia (1992)
8. Debreu, G.: Saddle point existence theorems. Cowles Commission Discussion Paper: Mathematicas No.412 (1952)
9. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**, 595–612 (2010)
10. Esfahani, P.M., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Math. Program. **171**, 115–166 (2018)
11. Guo, L., Lin, G.H.: Notes on some constraint qualifications for mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **156**, 600–616 (2013)
12. Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W.: Ambiguous joint chance constraints under mean and dispersion information. Oper. Res. **65**, 751–767 (2017)
13. Izmailov, A.F., Solodov, M.V.: An active-set Newton method for mathematical programs with complementarity constraints. SIAM J. Optim. **19**, 1003–1027 (2008)

14. Jin, C., Netrapalli, P., Jordan, M.: What is local optimality in nonconvex-nonconcave minimax optimization? In: International Conference on Machine Learning, pp. 4880–4889. PMLR (2020)
15. Kantorovich, L.V., Rubinshtein, S.: On a space of totally additive functions. Vestnik of the St. Petersburg University: Mathematics **13**, 52–59 (1958)
16. Lin, G.H., Chen, X., Fukushima, M.: Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. Math. Program. **116**, 343–368 (2009)
17. Lin, G.H., Fukushima, M.: Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey. Pac. J. Optim. **6**, 455–482 (2010)
18. Liu, Y., Pichler, A., Xu, H.: Discrete approximation and quantification in distributionally robust optimization. Math. Oper. Res. **44**, 19–37 (2019)
19. Luna, J.P., Sagastizábal, C., Solodov, M.: An approximation scheme for a class of risk-averse stochastic equilibrium problems. Math. Program. **157**, 451–481 (2016)
20. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
21. Mangasarian, O.L.: Error bounds for nondegenerate monotone linear complementarity problems. Math. Program. **48**, 437–445 (1990)
22. Mangasarian, O.L.: Global error bounds for monotone affine variational inequality problems. Linear Algebra Appl. **174**, 153–163 (1992)
23. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
24. Mangasarian, O.L., Pang, J.S.: Exact penalty for mathematical programs with linear complementarity constraints. Optimization **42**, 1–8 (1997)
25. Mangasarian, O.L., Ren, J.: New improved error bounds for the linear complementarity problem. Math. Program. **66**, 241–255 (1994)
26. Mangasarian, O.L., Shiau, T.H.: Error bounds for monotone linear complementarity problems. Math. Program. **36**, 81–89 (1986)
27. Nouiehed, M., Sanjabi, M., Huang, T., Lee, J.D., Razaviyayn, M.: Solving a class of non-convex min-max games using iterative first order methods. In: Advances in Neural Information Processing Systems, pp. 14934–14942 (2019)
28. Pang, J.S., Scutari, G.: Nonconvex games with side constraints. SIAM J. Optim. **21**, 1491–1522 (2011)
29. Pang, J.S., Su, C.L., Lee, Y.C.: A constructive approach to estimating pure characteristics demand models with pricing. Oper. Res. **63**, 639–659 (2015)
30. Pflug, G.C., Pichler, A.: Multistage Stochastic Optimization. Springer, Cham (2014)
31. Razaviyayn, M., Huang, T., Lu, S., Nouiehed, M., Sanjabi, M., Hong, M.: Nonconvex min-max optimization: Applications, challenges, and recent theoretical advances. IEEE Signal Process. Mag. **37**, 55–66 (2020)
32. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer-Verlag, Berlin (2009)
33. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)
34. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2014)
35. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optimization **57**, 395–418 (2008)
36. Su, C.L., Judd, K.L.: Constrained optimization approaches to estimation of structural models. Econometrica **80**, 2213–2230 (2012)
37. Villani, C.: Topics in Optimal Transportation. American Mathematical Soc., Providence (2003)
38. Xie, W.: On distributionally robust chance constrained programs with wasserstein distance. Math. Program. **186**, 115–155 (2021)
39. Xie, W., Ahmed, S.: On deterministic reformulations of distributionally robust joint chance constrained optimization problems. SIAM J. Optim. **28**, 1151–1182 (2018)
40. Xu, H., Liu, Y., Sun, H.: Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. Math. Program. **169**, 489–529 (2018)

41. Xu, H., Ye, J.: Approximating stationary points of stochastic mathematical programs with equilibrium constraints via sample averaging. Set-Valued Var. Anal. **19**, 283–309 (2011)
42. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J. Imaging Sci. **6**, 1758–1789 (2013)
43. Ye, J., Zhu, D., Zhu, Q.J.: Exact penalization and necessary optimality conditions for generalized bilevel programming problems. SIAM J. Optim. **7**, 481–507 (1997)
44. Ye, Y.: On affine scaling algorithms for nonconvex quadratic programming. Math. Program. **56**, 285–300 (1992)
45. Zhang, J., Xu, H., Zhang, L.: Quantitative stability analysis for distributionally robust optimization with moment constraints. SIAM J. Optim. **26**, 1855–1882 (2016)